

The Alberta Journal of Educational Research

Vol. XI, No. 1

March, 1965



THE COMMITTEE ON EDUCATIONAL RESEARCH

*Faculty of Education
University of Alberta*

A SURVEY OF THE LITERATURE RELATING TO PROBLEMS OF ADMISSION WITH PARTICULAR REFERENCE TO THE UNIVERSITY OF ALBERTA*

DONALD W. KNOWLES

A vital prediction problem faces university admissions officers today in attempting to solve the question, "Who should go to university?" A combination of a mounting tidal wave of students desiring to attend university and a pressing demand for the well-educated in virtually every field have emphasized the need for the efficient selection and classification of freshmen. The proportion of students in public school who eventually attend university increased from three per cent in 1890 to 8.7 per cent in 1957 to a projected ten per cent in 1965, according to the Dominion Bureau of Statistics (Woodside, 1958). This percentage increase is compounded by an increasing population. Sheffield has estimated that there will be a four-fold increase of university students within twenty years of 1955 (Woodside, 1958).

Canadian universities appear to be increasingly dissatisfied with the utilization of talented youth. In 1956, Dr. Andrew Stewart, then President of the University of Alberta, said:

Too many students who are admitted find it impossible, under conscientious direction of their instructors, to meet reasonable requirements for promotion . . . We do not know how many competent students fail to reach University, but the number is probably large (Jackson and Fleming, 1957; p. 80).

The essential problem has been to determine some method of predicting an individual's chances of success in university study. Some fifteen years ago, summaries of predictive research findings were compiled and evaluated by Travers (1949), Garrett (1949), and Henry (1950).

Travers (1949) found that the high school average was consistently the best single predictor of university success. Following in order of validity as predictors were subject matter tests and tests of scholastic aptitude. He also found that consistently low correlations were derived in studies of non-intellectual factors, such as level of adjustment and interests, probably due to inadequate measures of these factors.

Garrett's (1949) review considered hundreds of U.S. studies. He, too, found that the high school average had the greatest correlation with university scholastic average. The correlation coefficient of high school average with university average ranged from .29

*This is the first in a series of six related articles considering problems of predicting freshman success at the University of Alberta.

to .82 with a median of .56. He found that the high school average predicted the first-year university average better than it predicted any lesser or greater part of the university program. High school achievement was variously indicated by high school average and percentile rank-in-class. A combination of these two measures resulted in a median multiple correlation coefficient with university success of .65.

Henry (1950) also concluded that high school achievement had been shown to be the best single predictor. College aptitude tests were found to produce correlation coefficients ranging from .25 to .75 with college success, being fairly valid if devised for and used by a specific college. Personality measures failed to demonstrate any appreciable relationship with university success in most studies. Other factors such as size of family, nature of part-time employment, and parental occupation showed either low or inconsistent relationships.

Alberta Studies

Local studies have tested the research findings that measures of high school achievement, followed by standardized general achievement tests and intelligence tests, are the best predictors of university success.

Comprehensive Alberta studies in this field began in late 1954 with the establishment by the Department of Education of a Matriculation Study Subcommittee whose purpose it was to examine the "validity of the present predictive system of selecting students for university education" (Evenson and Smith, 1957; p. 68). The first project of the subcommittee was to examine the most recent, complete academic record of a university graduating class. The freshman class of 1951 was selected for study. The study was concerned with the predictive validity of such factors as the Grade XII Departmental results and the American Council on Education's (ACE) Psychological Examination (Evenson and Smith, 1958).

The results, reported in Table I, supported the finding of other studies that Grade XII achievement record is a better single predictor of university freshman success than an intelligence test (ACE). Further, the ACE consistently showed a smaller correlation with university average of each subsequent year of university program compared to the correlation derived by using Grade XII results. The 1951 Freshman Study also found that the best predictor of senior years' success in university was the freshman average (Table I). A major weakness of this study was that the group of subjects for one predictor variable was not the same as the group of subjects for any other variable. The number of subjects in the comparison groups ranged from 116 to 881, with a median size of

181 subjects. It would appear impossible to make precise comparisons between the correlation coefficients as, apparently, the size and nature of the comparison groups was not controlled.

TABLE I
CORRELATIONS BETWEEN UNIVERSITY AVERAGE AND
GRADE XII DEPARTMENTAL AVERAGE AND ACE TOTAL
SCORE FROM 1951 FRESHMAN STUDY
(After Evenson and Smith, 1958)

Criterion	Predictor*			Type of Student
	Grade XII Average	ACE	1st Year Univ. Ave.	
First Year University Average	.48	.36		All Faculties
	.58	.34		Arts and Science
	.59	.26		Engineering
	.54			Educ. (B.Ed.)
Second Year University Average	.54	.16	.63	All Faculties
Third Year (Grad.) University Av.	.36	.15	.45	All Faculties
Fourth Year (Grad.) University Av.	.50	.18	.47	All Faculties

*Comparison Groups Differ (Range of N, 116 to 881; N_{mdn} = 181)

A second project of the Matriculation Study Subcommittee was a study of the 1956 Grade XII graduating class in Alberta. Having found in the 1951 Freshman Study that the Grade XII Departmental mark was the best single predictor of first year university success, the Subcommittee investigated the relationship between Grade XII marks and other predictors. Specifically, the College Entrance Examination Board's (CEEB) achievement battery and the Scholastic Aptitude Test (SAT) as well as the School and College Ability Test (SCAT) were studied (Evenson and Smith, 1957).

The CEEB examinations have been used for many years by many universities in the U.S.A. as one basis for the selection of university students. The College Entrance Examination Board actually serves in an advisory and supervisory capacity to the Educational Testing Service which prepares the tests and conducts the examination program. In a study termed the "1956 Grade XII Survey," the Alberta Subcommittee chose seven of the CEEB achievement tests and SAT to administer to all Alberta Grade XII students.

A province-wide sample of 1,435 students was drawn. The students were assigned to 21 groups representing all possible combinations of two CEEB achievement tests from the battery of seven possible. Each student, then, wrote two of the CEEB tests corresponding to his Grade XII program. Correlations between the CEEB test results and the corresponding Departmental results were computed. The correlation coefficients ranged from 0.58 to 0.77, with a median correlation of 0.71 (Evenson and Smith, 1958). Despite the fact that the CEEB examinations were written for and normed on a U.S. population, they apparently correlate well with Alberta Grade XII Departmentals considering the differing types of instruments.

TABLE II

CORRELATIONS BETWEEN DEPARTMENTAL GRADE XII
SUBJECT AND CEEB ACHIEVEMENT TEST SCORES AND
FRESHMAN UNIVERSITY AVERAGE
(After Evenson and Smith, 1958)

CEEB Achievement Test	N _{CEEB}	Correlation with 1st Yr. Univ. Av.		N _{Dept}	Grade XII Dept. Score
		r _{CEEB}	r _{Dept}		
Adv. Math.	298	.41	.56	316	Math. 31
Interm. Math.	178	.34	.45	609	Math. 30
Physics	235	.38	.58	484	Physics 30
Chemistry	182	.38	.56	614	Chemistry 30
Biology	67	.26	.42	181	Biology 32
English	136	.20	.24	613	English 30
French	140	.29	.43	519	French 30
Sat — V	660	.27			
Sat — M	659	.30			

*Comparison Groups Differ

The next step of the 1956 Grade XII Survey was to test the validity of the data gathered in the special testing program to predict university success. Correlations between CEEB test scores and university first year average were compared with correlations between Departmental marks and university average for the 1956 freshman class (Table II). Although the significance of differences was not computed, due perhaps to the varying comparison groups, it appeared that Departmental marks consistently showed a higher

correlation with university average than did the CEEB examinations. Evenson and Smith (1958) concluded that there was no evidence that would support the CEEB's replacing the comparable Departmental examinations in selecting university students in Alberta.

The 1956 Grade XII Survey gathered other information such as scores on SCAT, responses to a Student Questionnaire on educational and vocational plans, and a principal's rating for each student on such factors as initiative, industry, extracurricular activities, interests in further education, and overall prospects of university success. Analyses have been conducted on only a small part of this data (Evenson and Smith, 1958). The correlation between SCAT, Level 1, and university first year average ($r=.29$) was very low and indicated that existing intelligence tests, as the ACE in the 1951 Study (Table 1), are of little value in predicting university success. Results of the incomplete analyses of the Student Questionnaire and principal's ratings on the nine factors showed these measures to be of limited value. (Evenson and Smith, 1958).

Black (1959, 1960) continued the analyses of data gathered in the 1956 Grade XII Survey. To compare Grade XII Departmentals with selected standardized tests used in the Subcommittee's study, 529 of the 1956 Grade XII class who took part in the special testing program and who entered the University of Alberta in 1956 were selected for study. The 1956 Study had provided scores for each student on two of seven CEEB tests, Scholastic Aptitude Test, Grade IX General Test, and SCAT, Level 1. The University Student Advisory Service provided scores on ACE Psychological Examination and two parts of the Co-operative English Test—Mechanics of Expression (Part A) and Reading Comprehension (Part C-2). As predictors of university first year average, these standardized tests ranked in the descending order of CEEB, SAT, SCAT, ACE, and Grade IX General Test (Black, 1959). No mention was made in this report of the relative effectiveness of the Co-operative English Test. Of the scholastic aptitude, or intelligence tests, the Scholastic Aptitude Test was most consistently the best predictor (Table III), although even its median correlation with university average was a modest .35 (Table III).

A disadvantage in the design of the 1956 Grade XII Survey was that each student wrote only two of the CEEB examinations with the result that comparison groups for each of the seven CEEB tests were different. Only limited comparisons could be made among the correlations derived by each of the CEEB tests with the university average. Black (1959) in comparing each of the CEEB tests with the corresponding Departmental examination, compared the

TABLE III
CORRELATION COEFFICIENTS BETWEEN FOUR ACADEMIC
APTITUDE TEST SCORES AND UNIVERSITY
FRESHMAN AVERAGE
(After Black, 1959)

		Sample writing part of CEEB Achievement Tests							Median r
		English	Biology	Chem- istry	Physics	Inter- mediate Math.	Ad- vanced Math.	French	
Test	N=	119	49	144	179	184	246	116	
General Test (IX)		.182	.234	.250	.227	.197	.237	.248	.234
ACE - L		.234	.414	.347	.294	.274	.338	.208	.294
ACE - Q		.292	.234	.090	.176	.162	.191	.194	.191
SCAT - V		.167	.197	.298	.322	.245	.285	.263	.263
SCAT - Q		.425	.315	.262	.301	.145	.253	.307	.301
SAT - L		.123	.291	.319	.271	.230	.296	.275	.275
SAT - M		.381	.376	.253	.293	.185	.349	.348	.348

TABLE IV
CORRELATIONS BETWEEN CEEB WITH UNIVERSITY
AVERAGE AND GRADE XII DEPARTMENTAL WITH
UNIVERSITY FRESHMAN AVERAGE
(After Black, 1959)

CEEB	N _{CEEB}	Correlation with Univ. Average		N _{Dept}	Grade XII Dept.
		r _{CEEB}	r _{Dept}		
Adv. Math.	246	.408	.572	246	Math. 31
Inter. Math.	184	.281	.300	184	Math. 30
Physics	179	.400	.568	179	Physics 30
Chemistry	144	.402	.563	144	Chemistry 30
Biology	49	.347	.552	49	Biology 32
English	119	.194	.168	119	English 30
French	116	.354	.470	116	French 30

coefficients derived for the same group of students for the two predictor variables. It was impossible, of course, to compare the CEEB average with the Departmental average for the same group of students since each student wrote only two CEEB tests. After comparing the respective correlations with university average for the individual CEEB tests and Departmentals for the same group of students (Table IV), Black (1959) found that Departmentals, with the exception of English 30, consistently did a better job of predicting university average than did the corresponding CEEB test. Since each of the comparison groups for a subject area were the same in Black's study, the derived correlation coefficients varied somewhat from those found by Evenson and Smith (Table II) for the same predictor variables. More assurance could be placed on Black's findings since he controlled many of the possible contaminators of correlation coefficients through his limited matching of groups of students.

Black (1960) and Zurowsky (1959) applied the technique of multiple correlation to compare combinations of the standardized tests used in the 1956 Grade XII Survey with combinations of Departmental marks. Black (1960) considered the criterion group as two groups—total university freshman sample ($N=529$) and Engineering freshmen ($N=131$). Zero-order correlation coefficients between the variables and university average (Table V) served to confirm the Departmentals' effectiveness as predictors. The three best single predictors, of those considered, were Departmental marks in Science, Mathematics, and Foreign Language for both of the criterion groups (Table V). The differentiation of the criterion groups appeared to be worthwhile as evidenced by the noticeable difference in correlation coefficients between them. No doubt, criterion groups which did not share membership (Engineering freshmen were included in the All Faculties group) would have led to even greater differences in the relative effectiveness of predictors.

Following the technique developed by Horst (1954), Black (1960) compared various combinations of predictor variables on the basis of multiple correlations. He found that a combination of the seven Grade XII Examination scores required for admission to Engineering yielded as great a multiple correlation coefficient with freshman Engineering average ($R_7=.686$) as the combination of twenty-five other predictor variables, the standardized tests described above and the Grade IX Departmental results ($R^{25}=.687$) (Black, 1960). Of the six combinations of predictor variables he studied, Black concluded that "the five predictor variables of Group B (English 30, Social Studies 30, Mathematics Average, Science Average, and Foreign Language Average) were deemed the best

operational combination for all the university course areas studied” (Black, 1960; p. 45). An important continuation of this study would seem to be to compare the predictive effectiveness of the Departmental results with that of a combination of Departmental results and a measure of scholastic aptitude, such as the Scholastic Aptitude Test. The incomplete CEEB scores for the students did not appear to be compensated for in this study so that any combinations of predictor variables which included CEEB scores may have been subject to contaminating factors arising from difference of comparative groups.

TABLE V
CORRELATION BETWEEN UNIVERSITY AVERAGE AND
SELECTED PREDICTOR VARIABLES FOR UNIVERSITY
FRESHMEN FROM 1956 GRADE XII SURVEY
(After Black, 1960)

Predictor Variable*	Correlation with University Average	
	All Faculties N = 529	Engineering N = 131
English 30	.227	.303
Social Studies 30	.425	.420
Gr. XII Sciences Av.	.632	.620
Gr. XII Math. Av.	.500	.598
Foreign Language	.434	.482
SCAT - Verbal	.234	.266
SCAT - Quantitative	.276	.254
Coop Eng. Mechanics	.235	.294
Coop Eng. Reading Comp.	.192	.271
CEEB Adv. Math.		.400
ACE - L		.361
ACE - Q		.245
SAT - V		.331
SAT - M		.336

*Partial list only.

TABLE VI
CORRELATIONS DERIVED FROM SEVERAL PREDICTOR
VARIABLES WITH INDIVIDUAL UNIVERSITY COURSE
MARKS AND FRESHMAN AVERAGE
(After Zurovsky, 1969; pp. 69-74)

University Course— Criterion	N	Zero-Order Correlations			Multiple Correlations			Best R minus Best r
		Median r	Best Single Predictor	Best r	Gr. XII Battery	SCAT COOP ENG.	Entire Battery*	
Chemistry 40, 42	336	.287	Sci. XII Av.	.691	.698	.360	.704	.013
Physics 40, 41	83	.260	Sci. XII Av.	.702	.698	.324	.698	— .004
Math. 1, 3, 4, 40, 45	137	.290	Sci. XII Av.	.621	.653	.338	.696	.075
Accounting 1	31	.311	Sci. XII Av.	.616	.655	.337	.741	.125
Commercial Law 41	31	.245	Soc. St. 30	.559	.604	.180	.713	.154
Botany 1, 5	65	.180	Sci. XII Av.	.720	.720	.105	.741	.021
Zoology 1	127	.220	Sci. XII Av.	.714	.715	.281	.717	.003
Geology 1	43	.159	Sci. XII Av.	.492	.526	.336	.670	.178
Geography 1	46	.305	Sci. XII Av.	.568	.608	.304	.713	.145
University Average	529	.255	Sci. XII Av.	.632	.656	.326	.658	.026

*Entire battery consisted of Grade IX Departmentals, Grade IX General Test, Grade IX Reading, Grade XII Departmentals, SCAT (V and Q), COOP English Tests.

The results of Zurowsky's (1959) study, which were included in the report made by Black (1960) were concerned with the relative predictive value of such variables as Grade IX Departmentals, Grade XII Departmentals, SCAT, Co-operative English Mechanics and Co-operative English Reading Comprehension with differential criteria—selected university science and business administration courses as well as university average. The subjects in this study were the 529 members of the 1956 Grade XII Survey who entered the university in 1956.

Against the criterion of university average, the largest zero-order correlation of predictor variables was found to be that of Grade XII Science average ($r = .632$). The multiple correlation between a combination of the Grade XII battery, SCAT (V and Q), and the Co-operative English tests with the criterion of university average was .656 (Zurowsky, 1959; pp. 69-71). The actual improvement in correlation coefficients by using the entire battery of predictor variables rather than the best individual predictor, which was the Grade XII Science average for 9 of 10 criteria, was apparently not tested for significance although the researcher intended to study, among other implications from past studies, that "the use of multiple correlation promised a means of selecting the best combination of variables for predicting each criterion" (Zurowsky, 1959; p. 46). The improvement, as indicated in Table VI, ranged from $-.004$ to $.178$ with a median improvement in correlation of $.050$. The generally slight improvement in correlation from using even all the predictor variables, as indicated in Table VI, would seem to provide some reason to doubt the implication above.

In comparing a battery of Grade XII Departmentals with a battery of SCAT subtests and Co-operative English tests combined, Zurowsky (1959) found that the Grade XII scores were the more effective predictors of university marks for nine individual courses and over-all average (Table VI). There was, generally, only slight improvement in the predictive ability of the entire battery of Grade IX and Grade XII Departmentals, SCAT, and Co-operative English tests as compared to the predictive ability of a battery of five Grade XII Departmentals alone, although, again, no results of testing for significance of differences were included in the report.

Zurowsky analyzed the contributions of each of the predictor variables in the Grade XII battery by comparing their beta weights in each multiple regression equation. He concluded that the Grade XII Science average consistently contributed more heavily than the other Grade XII subjects in predicting first year university success. The Grade XII Foreign Language mark did not contribute to the prediction batteries (Zurowsky, 1959; p. 93).

Clarke (1958) reported two other studies that had been conducted at the University of Alberta. Fitzpatrick studied 253 freshmen enrolled in the Engineering faculty in 1954-55. He found that the best predictor of university average was the Departmental mathematics average which yielded a correlation of 0.67. The ACE Psychological Examination yielded a correlation of only 0.23 with university average. Fair studied one hundred Arts freshmen who entered the university in 1953. He found that the correlation of the Departmental average with first year university average was 0.64; the corresponding correlation for the ACE Psychological Examination was 0.40. The relative superiority of Departmentals over the ACE was maintained in predicting the university third year average. The correlation coefficient using the Departmental average was 0.41, but the best predictor of third year average was the first year university average which yielded a coefficient of 0.69 (Clarke, 1958).

TABLE VII
CORRELATIONS BETWEEN SELECTED PREDICTOR
VARIABLES AND UNIVERSITY AVERAGE FOR 1960-61
UNIVERSITY FRESHMAN CLASS
(After Mack, 1963)

Group	Correlation with University Average					
	H. S. Average	ACE - Q	ACE - L	ACE - T	CO-OP, A	CO-OP, C ₂
Arts Faculty	.508	.155	.253	.260	.379	.321
Science Faculty	.715	.297	.282	.325	.384	.298
Education Fac.	.450	.084	.195	.174	.193	.219
Engineering Fac. ...	.349	.234	.243	.272	.151	.280
Other Faculties	.647	.130	.389	.354	.446	.390
MDN. r ...	.549	.155	.253	.272	.379	.298

A recent study by Mack (1963) of University of Alberta freshmen confirmed that the Grade XII Departmentals correlate more highly with university freshman average than do either the ACE Psychological Examinations or the Co-operative English tests. His results, related in part of Table VII, were computed in terms of the faculty in which a student enrolled. Such differential criteria allowed more precise probability tables to be drawn up in the second part of his study (Mack, 1963; p. 108). The Grade XII average correlated most highly with marks of Faculty of Science students

($r = .715$) and lowest with Education students ($r = .450$), varying from a good predictor in the first instance to a mediocre predictor in the second. Mack also grouped the high school mean scores into units of five in order to present a concise table of probabilities of success in each faculty. Such a technique, developed by Eells (1961-a) would appear to be a valuable form of presentation of predictive data for guidance purposes. Mack's suggestion that his tables could be used by high school students with their counsellors would seem to be limited by the nature of the data, Departmental marks, which are not available until after the school term has ended.

Implications Arising from Alberta Studies

1. Of the many types of predictors of university success considered in Alberta to date, the most valuable one has been the results of Grade XII Departmental Examinations. As shown in Table VIII, Grade XII marks have been found to be superior to standard achievement test scores, scholastic aptitude tests, and Grade IX results.

2. In studies of multiple prediction, the Grade XII battery has been found to be superior to batteries of other predictor variables. Very little improvement resulted from adding scholastic aptitude test scores or results from standard achievement tests to the Grade XII battery. Improvement in the prediction of success at the University of Alberta would appear to be found through improved methods of measuring achievement in Grade XII.

3. Either the Grade XII average or the Grade XII course marks were used in past studies. There was some indication that the use of course marks allowed the best discriminators, for example the science average, to be given more weighting than poorer predictors. There appeared to be a need to test the relative merits of Grade XII average as compared to Grade XII course marks in their ability to predict university success.

4. Recent studies in Alberta considered differential criterion groups at the university. Such differentiation appeared to offer the promise of improved predictive accuracy as compared to considering students as one total group.

5. As valid predictors are found, a problem of communicating the results to high school students, counsellors, and admissions officers has developed. The technique employed by Mack (1963) of using an expectancy chart would seem to be worthwhile. There was some indication that the regression equation appeared to be too complex a technique for current general school usage even though the graphic interpretation of multiple prediction equation scores proposed by Black (1960) offers a better interpretative technique for multiple-variable predictive data.

TABLE VIII
SUMMARY OF RESULTS OF ALBERTA STUDIES OF
CORRELATIONS BETWEEN UNIVERSITY AVERAGE
AND SELECTED PREDICTOR VARIABLES

Investigator	Type of Student	Predictor Variables				Type of Intel. Test
		Grade IX Dept.	Grade XII Dept.	Stand. Achiev.	Intel. Test	
Evenson and Smith (1958)	1951 Freshmen	.40	.48		.36	ACE-T
	1953 3rd Yr. Grad.	.42	.36		.16	ACE-T
	1954 4th Yr. Grad.	.25	.50		.18	ACE-T
Fair (Clarke, 1958)	1953 Arts Freshmen		.64		.40	ACE-T
	1955 Arts 3rd Yr.		.41		.31	ACE-T
Fitzpatrick (Clarke, 1958)	1954 Engin. Fresh. (N=253)		.58		.23	ACE-T
Evenson and Smith (1958)	1956 Freshmen		.45§	.30§†	.29	SCAT-T
Black (1959) (1960)	1956 Freshmen		.55§	.35§†		SCAT-Q
	1956 Freshmen		.43§	.24†	.28	ACE-L
	1956 Engineering Freshmen		.48§	.29†	.36	SCAT
Zurowsky (1959)	1956 Freshmen	.20§	.66*	.24†	.25§	ACE-T
Mack (1963)	1960 Arts Freshmen		.51	.38†	.26	ACE-T
	Science		.72	.38†	.33	ACE-T
	Education		.45	.19†	.17	ACE-T
	Engineering		.55	.15†	.27	ACE-T
	Other		.65	.45†	.35	ACE-T

§ Median correlation coefficient * Multiple correlation coefficient † CEEB Examinations ‡ Co-op. English Mechanics

REFERENCES

- Black, D. B. A comparison of the performance on selected standardized tests to that on Alberta Grade XII departmental examinations of a select group of University of Alberta freshmen. *Alberta J. Educ. Res.*, 1959, 5, 180-190.
- . The prediction of freshman success in the University of Alberta from Grade XII department results. *Alberta J. Educ. Res.*, 1960, 6, 38-53.
- Clarke, S. C. T. Review of some previous studies on matriculation problems. *Alberta J. Educ. Res.*, 1958, 4, 30-39.
- Eells, Kenneth. A vivid method of presenting chances for academic success. *J. Counsel. Psychol.*, 1961—a, 8, 344-350.
- . How effective is differential prediction in three types of college curriculum? *Educ. Psychol. Measmt.*, 1961-b, 21, 459-471.
- Evenson, A. B. and Smith, D. E. A study of matriculation in Alberta. *Alberta J. Educ. Res.*, 1957, 3, 62-71.
- . A study of matriculation in Alberta. *Alberta J. Educ. Res.*, 1958, 4, 67-83.
- Garrett, Harley F. A review and interpretation of investigations of factors related to scholastic success in colleges of arts and science and teachers colleges. *J. Exp. Educ.*, 1949, 18, 91-138.
- Henry, Edwin R. Predicting success in college and university. In D. H. Fryer and E. R. Henry (Ed.), *Handbook of Applied Psychology*. Vol. II, Toronto: Rinehart and Co., Inc., 1950.
- Horst, Paul. A technique for the development of a differential prediction battery. *Psychol. Monogr.*, 1954, Vol. 68: No. 9, Whole No. 380.
- Jackson, R. S. B. and Fleming, W. G. Who goes to university? In C. T. Bissell (Ed.), *Canada's Crisis in Higher Education*. Toronto: Univer. of Toronto Press, 1957.
- Mack, L. L. Examining the efficiency of predictors presently being used at the University of Alberta. *Alberta J. Educ. Res.*, 1963, 9, 100-110.
- Travers, Robert M. W. Significant research on the prediction of academic success. In Donahue et. al. (Ed.), *The Measurement of Student Adjustment and Achievement*. Ann Arbor: Univer. of Michigan Press, 1949.
- Woodside, Willson. *The University Question*. Toronto: Ryerson Press, 1958.
- Zurowsky, John. Predicting freshman success in seven science and two business administration course at the University of Alberta. Unpublished master's thesis, Univer. of Alberta, 1959.

THE PORTEUS MAZE TEST AND MANUAL PROFICIENCY

BARRY P. FROST*

University of Alberta, Calgary

The Problem

One of the problems in the education of mentally deficient adolescents is the selection of appropriate manual tasks on which to train them or vice versa, the direction of different groups of defectives to the tasks which are suitable for them.

The Christine Meikle School for Retarded Children, in Calgary, has an excellent workshop in which retarded males construct various articles from wood. One of the difficulties faced by the staff is that of deciding which boys are ready for work on the more complex apparatus such as band saw and lathe. In this situation the staff need a selection technique which can be simply and quickly administered. The Porteus Maze Test was considered by the writer to be such a technique.

There have not been many studies using this technique in a manual training or industrial setting. Dr. Henry Herd, medical officer to the Manchester (United Kingdom) special schools in 1930 obtained manual ability ratings from the teachers of 130 children. Those children rated "A" in ability had a mean maze test quotient of 77, "B" of 74, "C" of 65, "D" of 59, while those rated lowest averaged 46. He stated that "no child with a Porteus ratio below 50 made anything but a very poor appearance in manual work." (Herd, 1930)

Earle and Milner (1929) reported that the achievement of apprentice fitters, electricians and smiths, at the end of the training period, correlated 0.61 with the Porteus Maze scores.

Jensen (1952) reports on the usefulness of the Maze test in the selection of Air Force trainees. He writes that "every individual scoring high in the Porteus Test presents qualitative evidence of power to plan ahead, to execute with prevision and to adapt readily to tasks of increasing complexity."

Method

Twenty-six male retardates, ranging in age from thirteen to twenty-two years, currently enrolled in the Christine Meikle School

*The writer would like to thank Miss Darlene Velker, who carried out the testing, Mr. Roy Bray, who was responsible for the workshop ratings, Mrs. D. L. Hood who was responsible for the academic ratings, and Mrs. Christine Meikle, principal of the Christine Meikle School for Retarded Children, for their help in carrying out the project.

for Retarded Children, were rated by the teacher in charge of the workshop, on a ten point scale, for their ability to perform wood-work. Another teacher independently rated them on a ten point scale for their academic work.

Independently of the two teachers, and without knowing any details concerning the subjects, a post graduate student, under the supervision of the writer, administered the Porteus Maze Test. The administration was carried out individually in a quiet room. Thus the testing and rating were completely uninfluenced by each other.

The Maze Test was scored by taking as the score the year level of the last maze successfully completed. This score was then correlated with the two ratings which were also correlated with each other (Pearson Product-Moment Method).

Results

The major results are as shown in Tables I, II, and III.

The maze test correlates 0.65 with the rating of the retardates on their past performance in the wood workshop. Though only a moderate correlation it would seem to be large enough to encourage the trial use of the test as a selection device.

TABLE I
DISTRIBUTION OF WOODWORK RATINGS
ACCORDING TO MAZE SCORE

Maze Score	Woodwork Ratings
3	1, 1, 1
4	4
5	2, 9, 4, 3, 3, 2, 2
6	3
7	5
8	1, 4, 4, 2, 3, 2
9	6, 8
10	8, 6
11	10,
12	10, 5

TABLE II
RESULTS OF USING CUT-OFF POINTS OF MAZE SCORE 9
AND WOODWORK RATING 6

		Woodwork Rating		
		6+	5—	Σ
Maze Score	9+	6	1	7
	8—	1	18	19
	Σ	7	19	26

TABLE III
PRODUCT-MOMENT CORRELATION COEFFICIENTS

	Woodwork Rating	Academic Rating
Maze Test	0.65	0.39
Woodwork Rating		0.36

By inspection of Table I it may be seen that an arbitrary cut-off point at Score 9 would select those retardates who have definitely shown their capacity to handle the workshop machinery and techniques.

There is a slight tendency for the boys with high woodwork ratings to be older than those with low ratings. Those boys (N = 7) who obtained a rating of 6 or higher have a mean age of 18.72 years (S.D. 2.07), whereas the boys with a rating of 5 or less have a mean age of 16.95 years (S.D. 2.99). But if we take the sample as a whole the woodwork rating scales correlate 0.01 with age.

There is virtually no age difference between those scoring high (9+) on the Maze Test and those scoring low (8—). The mean ages and S.D.S. are 17.28 (2.74) and 17.47 (2.98) respectively. Over the whole sample the Maze Test correlates —0.10 with age. As there appears to be virtually no correlation with age in this retarded group, there is no reason for converting the maze ages to quotients.

It is obvious, by inspection of the correlation co-efficients that aptitude at woodwork is relatively independent of aptitude at academic work and that the Maze Test is a better predictor of the former. The use of academic ratings in a multiple correlation together with the Maze Test Scores is of no value as the resultant R is 0.66 as against the r value of 0.65 for the Maze Test Scores alone.

Conclusions

1. The Porteus Maze Test has reasonable promise as a simple technique for selecting retardates for advanced work in the woodwork shop.

2. Success on the Maze at Year 9 seems to be the minimum requirement for prediction of adequate performance in woodwork.

REFERENCES

- Earle, F. M., and Milner, M. et al. *The Use of Performance Tests of Intelligence in Vocational Guidance*. London, H.M.S.O. 1929. (Quoted in 4)
- Herd, H. *The Diagnosis of Mental Deficiency*. London, Holder and Stoughton 1930. (Quoted in 4)
- Jensen, M. B. "Basic Airmen Who Fail to Meet Minimum Mental Standards for Retention in the Air Force." Mimeographed report Nov. 1952. (Quoted in 4)
- Porteus, S. D. *The Maze Test and Clinical Psychology*. Palo Alto, Pacific Books, 1959.

A COMPARISON OF READING ACHIEVEMENT OF
CHRISTCHURCH, NEW ZEALAND AND EDMONTON
ALBERTA PUBLIC SCHOOL STUDENTS OF THE SAME
AGE AND NUMBER OF YEARS OF SCHOOLING

by

MARION E. FOSTER,

Reading Specialist, Edmonton Public School Board

and

DONALD B. BLACK

Faculty of Education, University of Alberta

Comparisons of performance of one group to another are at the same time both enticing and potentially misleading. The dangers lie not so much in the actual comparisons but in the way they are made, and more important, the way the findings are interpreted. Perhaps this explains why such educational comparisons when made rarely appear in the literature notwithstanding their obvious usefulness to the educational systems concerned. When two systems are at opposite sides of the globe, the dangers of misinterpretation due to differences in educational philosophies, school organization, and curricula are that much greater. However, the subject of this investigation, reading, is one which in two areas of the Commonwealth whose populations are predominately English-speaking, can be truly viewed as the common medium through which the bulk of the schools' efforts is communicated. Therefore, regardless of differences in philosophy and/or methodology, actual reading performance becomes crucial. When differences in reading performance are found to exist, reasons explaining such differences should not alone be sought, but hopefully an examination of means effecting improvement will be made as well. It is obvious, too, that all in education is not black nor white, but rather shades of grey, and thus each party may learn from the other.

A frequent limitation of many comparative studies is that an instrument is chosen that favors one group or another. Such was the case in an earlier study by the authors (Black and Foster, 1963). Recently, there has been some effort through the offices of UNESCO, notably in the fields of mathematics, to develop evaluative instruments so equated as to be suitable for all populations. Such instruments were not available for this study, but an effort was made to circumvent this problem by administering to both groups two instruments which were characterized by one, in turn, being more familiar to one group than to the other. In this study, one instrument used was the Australian Council for Educational Research

(A.C.E.R.) a set of Silent Reading Tests which was standardized for New Zealand schools by the New Zealand Council for Educational Research. The other instrument, familiar in Alberta, was the California Reading Test (CRT).

The design of the study as noted was to test each group with both instruments after first establishing the comparability of academic aptitude, again by using a test recommended and used by the New Zealand Council for Educational Research, in this study, the Otis Quick-Scoring Mental Ability Tests (Beta, Form CM), and the California Short-Form Test of Mental Maturity (Elementary, 1957 S-Form) which is frequently used in Alberta. The California Test Bureau instrument, hereafter called the CTMM, gives both a Language and Non-Language score as well as a Total score; The Otis, a single score. The California Reading Test (Junior High Level, Form W), gives two scores: Reading Vocabulary, and Reading Comprehension as well as a Total Reading score. Three A.C.E.R. tests were used: Word Knowledge (Form C), Speed of Reading (Form B), and Reading for Meaning (Form C). Inasmuch as the same forms were used for all samples all statistical analyses were conducted using raw scores with one exception. All Otis scores were calculated in I.Q. scores using the approved New Zealand norms.

As has been noted in the earlier study in arithmetic by the authors, (1963) a problem was encountered in the selection of comparable samples. New Zealand children start school a year earlier than their Alberta counterparts. This problem was overcome by choosing two Christchurch samples, one group, Standard Four's, had attended school the same number of years as the Edmonton sample, but were on the average a year younger; and a second sample, Form I's, was the same age but had had an additional year of schooling. The two Christchurch samples were from five schools selected by the Senior Inspector of Schools in Christchurch as being representatives of the Christchurch Standard Four and Form I populations. The Edmonton sample was from classes in ten schools selected at random from all Edmonton schools with Grade Six classes with the exception of six schools which were conducting experimental programs in homogeneous grouping at the time of the study. In all, the sample consisted of 476 Grade Six Edmonton pupils, and 241 Christchurch Standard Four and 272 Form I pupils. These samples represent all pupils for whom scores on all variables used in this study were reported.

The means and standard deviations for each sample by sex and for the total sample were calculated and are reported in Table I. The conventional t-test for the significance of differences of means

were conducted and the findings reported in Table II. The F-Test was used to test for the homogeneity of variances and these results are reported in Table III. Throughout the null hypothesis was used, namely, that the differences in performances found could be attributed to the operation of chance factors when the comparisons made were between Edmonton and Christchurch pupils. However, when significant differences were found, the confidence level of this difference was noted and the sample with the largest statistic identified. In the comparisons of the two Christchurch samples, it was hypothesized that Form I students, as a group, would perform better than the Standard Fours on all variables but I.Q. Age was reported in months and based on the beginning of the school year in each city. It was hypothesized that by the very method of selecting the sample the Standard Four students would be twelve months younger on the average than either the Form I or Grade Six pupils who would be the same age. The findings may be summarized as follows:

1. The mean ages of the samples confirm the intended design of the study. However, Christchurch students tended to be less variable in age than did the Edmonton sample. The ages of the Christchurch Form I girls tended to be more variable than the Standard Four girls at the 5% level of confidence.
2. On the Otis, Edmonton Grade Six had higher I.Q.'s than did both Christchurch samples at the 1% level of confidence, although there were no significant differences in the variances of the three groups' I.Q.'s with the single exception of the Christchurch Form I and Edmonton Grade Six boys, the latter sample being more variable.
3. The raw scores on the three CTMM scores show significant differences favoring the Edmonton sample at the 1% level of confidence in all instances but one, namely, that of Non-Language scores between Edmonton Grade Six girls and Christchurch Form I girls when no statistically significant differences were found. Similarly, Christchurch Form I's as a group performed significantly better at the 1% level of confidence with the sole exception of the boys on the Non-Language Test where no statistically significant differences were found. In terms of the variability of scores, there tended to be no significant differences found in the Language scores with the exception of Edmonton Grade Six Girls who were more variable than Christchurch Form I girls, and Christchurch Standard Four girls, whose scores were more variable than Form I girls. On the Non-Language scores, the Edmonton sample was less variable at the 1% level of confidence than either of the two Christchurch samples in all instances.

TABLE I: MEANS AND STANDARD DEVIATIONS BY SEX OF READING, INTELLIGENCE, AND AGE VARIABLES FOR THE THREE MAJOR SAMPLES

Population	Sample		Statis. IC	Age Month	OTIS (IQ*) Total	G. T. M. M.		A. C. E. R.			C. R. T.			
	Sex	No.				Lang- uage	Non- Lang.	Total	Word Know.	Speed Read.	Read. Mean.	Read. Vocab.	Read. Comp.	Total
Edmonton Grade VI**	Female	231	$\bar{X}$	144.19	110.25	51.68	48.12	99.80	50.84	11.16	42.43	37.06	47.74	84.80
			S.D.	8.3	11.7	9.1	6.8	13.7	13.0	2.9	9.1	9.9	11.0	19.1
	Male	245	$\bar{X}$	145.46	108.48	51.30	48.94	100.24	48.70	9.82	40.16	35.77	47.14	83.18
			S.D.	8.3	14.0	9.3	7.0	13.9	14.1	3.7	9.5	10.4	12.2	20.9
	Total	476	$\bar{X}$	144.85	109.34	51.49	48.54	100.04	49.74	10.47	41.27	36.40	47.57	83.97
			S.D.	8.3	13.0	9.2	6.9	13.7	13.6	3.4	9.3	10.1	11.6	20.0
Christchurch Standard Four **	Female	113	$\bar{X}$	132.38	104.83	42.73	40.18	82.92	38.54	11.06	35.73	25.80	33.34	59.15
			S.D.	4.2	12.0	8.9	10.7	16.6	15.3	3.9	10.3	10.9	11.9	20.8
	Male	128	$\bar{X}$	134.17	102.14	43.88	40.10	83.99	39.42	9.80	32.16	25.99	34.21	60.21
			S.D.	6.2	11.9	8.9	10.8	16.8	16.0	3.6	10.7	11.2	12.3	21.9
	Total	241	$\bar{X}$	133.34	103.40	43.34	40.15	83.49	39.01	10.39	33.84	25.90	33.81	59.71
			S.D.	5.2	12.1	8.9	10.7	16.7	15.7	3.9	10.7	11.0	12.1	21.4
Christchurch Form I**	Female	127	$\bar{X}$	144.64	105.31	47.22	46.98	94.13	49.68	14.59	38.96	33.09	41.91	75.03
			S.D.	5.1	11.4	6.8	9.2	14.0	13.5	4.4	9.0	10.3	12.2	20.8
	Male	145	$\bar{X}$	146.66	101.93	45.35	45.51	90.98	46.40	11.84	36.52	30.66	39.50	70.25
			S.D.	6.7	12.7	9.0	10.3	15.6	15.8	4.2	11.1	11.5	11.0	20.5
	Total	272	$\bar{X}$	145.72	103.51	46.23	46.20	92.46	47.94	13.13	37.66	31.80	40.63	72.49
			S.D.	6.1	12.2	8.1	9.8	14.9	14.9	4.5	10.2	11.0	11.7	20.8
TOTAL POSSIBLE SCORE ON TEST					--	80	60	140	100	31	60	60	86	146

*I.Q. are calculated from Norms provided for New Zealand. All other statistics reported are raw scores.
**Subsequent tables will identify Edmonton Grade Six students as Edmonton 6; Christchurch Standard Four's, as Christchurch 6; and, Christchurch Form I's as Christchurch 7. This coding is based on years of schooling.

TABLE II: LEVELS OF CONFIDENCE OF REPORTED MEAN DIFFERENCES IN PERFORMANCE ON READING, INTELLIGENCE AND AGE VARIABLES BY SEX FOR THE THREE MAJOR SAMPLES

Sample Comparison		Age (Months)	Otis (IQ) Total	C. T. M. M.			A. C. E. R.				C. R. T.	
				Lang- uage	Non- Lang.	Total	Word Know.	Speed Read.	Read. Mean.	Read. Vocab.	Read. Comp.	Total
Edmonton 6 Versus Christchurch 6*	Female	1%	1%	1%	1%	1%	1%	N. S. D.	1%	1%	1%	1%
	Male	1%	1%	1%	1%	1%	1%	N. S. D.	1%	1%	1%	1%
	Total	1%	1%	1%	1%	1%	1%	N. S. D.	1%	1%	1%	1%
Edmonton 6 Versus Christchurch 7*	Female	N. S. D.	1%	1%	N. S. D.	1%	N. S. D.	1% c*	1%	1%	1%	1%
	Male	N. S. D.	1%	1%	1%	1%	N. S. D.	1% c*	1%	1%	1%	1%
	Total	N. S. D.	1%	1%	1%	1%	N. S. D.	1% c*	1%	1%	1%	1%
Christchurch 7 Versus Christchurch 6**	Female	1%	N. S. D.	1%	1%	1%	1%	1%	1%	1%	1%	1%
	Male	1%	N. S. D.	N. S. D.	1%	1%	1%	1%	1%	1%	1%	1%
	Total	1%	N. S. D.	1%	1%	1%	1%	1%	1%	1%	1%	1%

*Although the hypothesis has stated that there will be no significant differences between Edmonton and Christchurch, the levels of confidence of differences favoring Edmonton students are reported with the exceptions favoring the Christchurch sampled noted as 1% c.

**The significance of differences all favor Christchurch 7 (Form I) except as noted.

TABLE III: LEVELS OF CONFIDENCE OF REPORTED DIFFERENCES IN VARIANCES OF PERFORMANCE ON READING, INTELLIGENCE AND AGE VARIABLES BY SEX FOR THE THREE MAJOR SAMPLES

Sample Comparison		Age (Months)	Otis (IQ) Total	C.T.M.M.		Total	A.C.E.R.		C.R.T.			
				Lang- uage	Non- Lang.		Word Know.	Speed Read.	Read. Mean.	Read. Vocab.	Read. Comp.	Total
Edmonton 6 Versus Christchurch 6*	Female	1%	N.S.D.	N.S.D.	1%*	1%*	5%*	1.%*	N.S.D.	N.S.D.	N.S.D.	N.S.D.
	Male	1%	5%	N.S.D.	1%*	5%*	N.S.D.	N.S.D.	N.S.D.	N.S.D.	N.S.D.	N.S.D.
	Total	1%	N.S.D.	N.S.D.	1%*	1%*	1%*	N.S.D.	N.S.D.	N.S.D.	N.S.D.	N.S.D.
Edmonton 6 Versus Christchurch 7*	Female	1%	N.S.D.	1%	1%*	N.S.D.	N.S.D.	1%*	N.S.D.	N.S.D.	N.S.D.	N.S.D.
	Male	1%	N.S.D.	N.S.D.	1%*	N.S.D.	N.S.D.	N.S.D.	5%*	N.S.D.	N.S.D.	N.S.D.
	Total	1%	N.S.D.	5%	1%*	N.S.D.	N.S.D.	1%*	N.S.D.	N.S.D.	N.S.D.	N.S.D.
Christchurch 7 Versus Christchurch 6*	Female	5%	N.S.D.	1%*	N.S.D.	5%*	N.S.D.	N.S.D.	N.S.D.	N.S.D.	N.S.D.	N.S.D.
	Male	N.S.D.	N.S.D.	N.S.D.	N.S.D.	N.S.D.	N.S.D.	N.S.D.	N.S.D.	N.S.D.	N.S.D.	N.S.D.
	Total	N.S.D.	N.S.D.	N.S.D.	N.S.D.	N.S.D.	N.S.D.	5%	N.S.D.	N.S.D.	N.S.D.	N.S.D.

* Asterist notes that when a significant difference in variance is found it is the second named of the two samples compared that has a significantly larger variance, otherwise, it is the first named.

4. On the six variables from the Reading performance tests, the Edmonton Grade Six pupils' mean performance was significantly higher, at the 1% level of confidence, than both of the Christchurch groups with three exceptions. There were no significant differences between the Edmonton sample and the Christchurch Standard Four's on Speed of Reading and with the Christchurch Form I's on Word Knowledge. The Christchurch Form I's as a group had statistically significant higher mean performance than the Edmonton sample on the Speed of Reading Test. In all instances, the mean performance of the Christchurch Form I's was higher than that of the Standard Four's.

Significant differences in the variance of the group scores were found between Edmonton girls and Christchurch Standard Four girls on Word Knowledge (5%) and Speed of Reading (1%) and with Form I girls on Speed of Reading (5%). In all instances, the variability of the Christchurch sample was greater. Christchurch Form I boys showed significantly greater variability in the Reading for Meaning Test than Edmonton boys at the 5% level of confidence. These differences, in part, are reflected in the differences in variability found for the Total sample, i.e., Edmonton and Christchurch Standard Four's on Word Knowledge at the 1% level, with Form I's on Speed of Reading at the 1% level. No total variance differences were found to be significant for Speed of Reading for Edmonton and Standard Four samples, and for Edmonton and Form I samples Reading for Meaning Test scores. A larger variance favoring the Form I sample was found to be significant at the 5% level of confidence than that found for the Standard Four sample on the Speed of Reading Test.

Conclusions and Discussion

There is a marked absence of available research of the type reported here. Much of what is to follow will depend upon the subjective opinions of the investigators. New Zealand and Alberta have educational histories which are marked by their similarity of evolution. As Borland (1959) points out, the New Zealand schools were influenced in turn by the English and Scottish schools and now by the American approach. Certainly, the remnants of the earlier development is not as evident in Alberta as New Zealand, for example, while Alberta readers clearly show an American influence, Christchurch teachers have a choice of English or American readers. Concern has been expressed that both types of readers do not meet necessary conditions of New Zealand culture and the experience domains of the pupils, for example, snow plows and deep snow. Efforts have been underway to develop a unique set of readers for New Zealand schools with their publication expected

shortly. To meet this problem, much use is made of a Departmental publication "The School Reader," which has a high 'local' content. Another interesting and perhaps significant difference is the absence of controlled vocabularies in all New Zealand reading materials. Evidence further suggests New Zealand school libraries fare even less well than those in Alberta which themselves fall far below the level thought necessary. Certainly, one clear cut difference between the two systems lies in the New Zealand emphasis on oral expression, an aspect of education all but neglected in Alberta schools. The effects of this program emphasis are readily apparent both in the classroom and in the speech of the New Zealand general public.

The qualifications of the Christchurch teachers and their methods of teaching reading are very similar to those for Edmonton teachers at the level of the study. Similarly, there is an equal proportion of speech and reading problems which require special and individual attention. However, unlike Edmonton where class enrollments are in the thirties, in Christchurch, they are in the fifties. The time available for individual attention to reading difficulties, therefore, is proportionately lowered to that found in Edmonton and may explain in part the larger variability of performance in the Christchurch samples, and possibly, one reason for a lower overall performance.

As noted in the experimental design of this study, the majority of children start school in Christchurch at age five, or a year earlier than their Edmonton counterparts. However, it is noted that the variability of ages of the Edmonton sample is greater than that of the two Christchurch samples. This may well be attributed to a more rigorous adherence to the practice of "social promotion" than occurs in Alberta. This may account, as well, for the smaller variances of ages of the two New Zealand samples and, in part, for the generally larger range of performances on the performance tests. In only six of thirty performance comparisons were Christchurch standard deviations numerically greater than those in Edmonton. It may well be, however, that this is more directly a result of class size and its consequences when it is noted that in contrast to the large foreign born or direct foreign descent from culturally isolated areas in Northern Alberta that is evident in Edmonton, ninety percent of the New Zealand population are native born and of direct British descent, with only two percent of the population born in foreign lands. Under such circumstances, it could well be expected that the variability of scores in Alberta schools would have been greater rather than the reverse situation found in this study.

Cultural differences in the instruments used to measure performance must be recognized. For example, the CTMM is clearly

recognized as an American test. Similarly, the Otis test, although approved for New Zealand schools is still an American test. Certainly to the extent that these tests are in fact culturally biased, the advantage must be with the Alberta sample. Familiarity with the instrument's content form or this method of assessment will also be influential in determining performance. However, there is little reason to believe that the New Zealand students were unfamiliar with most of the vocabulary of the instruments or that by elimination in such instances when they were not familiar with terms, they could not arrive at the correct response. For example, it should be recognized that while petrol is the common term for gasoline in New Zealand, the question which asks which of the following is a fuel, gasoline, motorcycle, smoke, wheels or horn, the latter four can be quickly and obviously eliminated. More serious is the content of the two intelligence tests which involve arithmetic and science items. The problems of mathematical items, notably, those involving money and the decimal system have been referred to in the earlier article by the authors. Some items on the tests do involve science content, a subject which is delayed until Form I in New Zealand schools. On the other hand, it must be recognized as well that reading ability does correlate highly with intelligence tests scores so that it is possible that if one group or the other are, in fact, poorer readers, this will be reflected in lower scores on the intelligence instruments, and the cultural novelty of the item content will only serve to exaggerate this effect.

In terms of the content of the instruments measuring reading performance, it is noted that even in the A.C.E.R. instruments there are words, particularly in the vocabulary test, with which Edmonton children would have an advantage, for example, 'escalator' and 'supersonic'. However, the excellent speaking background of the majority of children in New Zealand schools who have the further advantage of a fine and broad choice of words heard in everyday speech should tend to favor them. No differences were found between the Edmonton and Christchurch samples of like age on the A.C.E.R. vocabulary test which suggests that vocabulary may be related more closely to maturation than years of formal schooling. On the vocabulary sub-test of the California Reading Test, Edmonton students were found to be superior, but as noted earlier in the discussion of the CTMM scores, this could be related to the delay in teaching science and fractions because the CRT does contain such items in this section.

Edmonton students did not fare as well on the Speed of Reading Test, being equalled by the Standard Four's and exceeded by the Form I's. This suggests that training is more influential in determining performance in this area than maturation, the children of

like age in New Zealand having had more instruction and practice in reading. Further, this could be assisted by the daily experience of each pupil with correct grammar and choice of precise words which characterizes New Zealand speech. Thus, the exact language of the printed page would appear more familiar to them. This stands in contrast to the poor grammar and limited speaking vocabulary heard by a great number of Alberta children. It may well be, and this is also suggested in part by the research on reading ability, that because of the emphasis on oral expression which obviously results in the spoken fluency so apparent in New Zealand children, this could also account for the smooth oral reading of the children which, in turn, could greatly influence smooth, rapid, silent reading. Further support for this hypothesis is found in the observation by one of the writers that indicated that New Zealand children did not exhibit weaknesses of finger movement and lip reading to the same extent as had been observed in Alberta children.

Under conditions of equality of vocabulary and with the advantage of speed of reading found in the study, it becomes difficult to account for the findings that the Edmonton sample had greater comprehension of the material read than did either Christchurch sample on both reading instruments. A possible explanation may lie in the content of the items of both these instruments as it relates to what and when, curricular material, such as science, social studies, is either taught or emphasized in Christchurch schools. The delay in the study of foreign countries and of chart and graph making might have put the New Zealand sample at a disadvantage as to terms and thought patterns used in at least two sections of the CRT. Certainly, the method of teaching social studies as observed in New Zealand schools does differ markedly from Alberta schools where with easy access to a variety of source materials and reference books students are expected to select information from factual excerpts. Further, unlike North Americans, New Zealanders do not show the same enthusiasm for statistics in their daily lives. This is most apparent in the comparison of the content of the A.C.E.R. and CRT comprehension tests. While the former asked for the literal idea, the main idea, or the best title of the story, the CRT may have overwhelmed the Christchurch students by asking them from many facts and figures, to perceive sequence, find the main theme or trend, observe relationships, classify, summarize and draw conclusions.

The evidence remains that, on the whole, the Christchurch samples did not perform as well on both comprehensive measures. It could be that a limited use of tests standardized for either local or national (or international) use has not provided New Zealand teachers with a standard against which to judge pupil performance.

It must be recognized that the testing protocols as found in the instruments are not as widely used in New Zealand classrooms as in Edmonton classrooms. It must be commented here that even in Edmonton classrooms the use of such testing protocols is not widespread, but in relative degree, the advantage must lie with the Edmonton sample. Certainly, the content of the items more closely parallels the type of things that the Edmonton teacher does in the classroom but not necessarily tests formally.

In conclusion, this study has left the investigators with more questions than when they began. Problems of cultural (curricular) influence in testing is not an unknown problem. The matter of test item protocol on response set is just beginning to be intensively studied. While the New Zealand sample did not perform as well as the Edmonton samples on the criterion instruments, it may well be that the methods used result, as a survey has shown, in New Zealanders' reading more books per capita than Canadians (Libraries, 1960). It could well be that the recognized slower pace of life in New Zealand creates a response set contrary to the competitive set that so often characterizes North Americans. Clearly, there is a need for far greater research of a comparative nature which will help educators understand and appreciate the advantage of other systems for our own systems that we may all profit from our experiences. For example, experimentation in Alberta with increased teaching of oral expression may both explain and produce the advantages in reading found in the New Zealand sample.

REFERENCES

- Black, D. B. and M. E. Foster. "A Comparative Study of the Performance in Arithmetic of Edmonton, Alberta, and Christchurch, New Zealand, Public School Pupils of Comparable Age and Grade Level." *Alberta Journal of Educational Research*, IX: 1 (March 1963), 49-59.
- Borland, Ron. "Time for Our Educationalists to Think Where We Are Going." *National Education*, XLI: 449 (November, 1949), 4455.
- "Libraries." New Zealand Official Year-Book 1960. Wellington, N.Z. R. E. Owen, Government Printer, 1960.

SPELLING SKILL AND PREFERENCE IN CENTRAL ALBERTA

ARTHUR G. STOREY

Faculty of Education, University of Alberta, Calgary
and

LORNE W. BUNYAN

Principal, Calgary Separate Schools

Rationale

From Grades II through VIII, spelling forms a significant part of the Alberta school curriculum. Some formal attention is generally paid to it in Grade IX also, but from this point forward, however, growth in spelling skill is left pretty much to the whims of the individual teachers and students. To the end of the Junior High School level, then, a good deal is known concerning spelling growth. Interest in the subject and knowledge about it wanes from Grade IX on, yet persistent questions and rumors as to spelling skill and preference continue to pervade the higher educational scene.

The present study was designed to answer questions like: 1. Does growth in spelling continue beyond the Junior High School level and if so where does it reach its maximum? 2. Do Albertans favor American (1), (2) or British (3) spelling? 3. Is there a shift in spelling preference as students proceed through higher academic levels owing to the preponderance of American textbooks and other printed matter used in the province's institutions of higher learning? In seeking answers to these questions the present authors surveyed a large number of students and graduates in the central Alberta area. In this survey the respondents were asked to choose a spelling preference for each of seventeen words selected on the bases of four criteria: 1. common usage, 2. difficulty from a spelling standpoint, and 3. clear and popular British and American alternatives, and 4. a common and logical misspelling alternative. Table I gives the words used in the survey together with the spelling alternatives. The words were presented to the subjects in the random order in which they appear in Table I but without the words being identified as indicated in the legend.

Sample

The subjects chosen for the survey included all of the rural Grade X students in the Calgary Inspectorate, that is all of those students taking Grade X at schools situated between two lines drawn across Alberta at High River and Olds, but exclusive of those in the city of Calgary. It was the authors' intention to use this large

sample as a basis against which to consider subjects at various higher levels of academic training. The second group in the sample included all of the available out-of-school Grade XII graduates in the same area but inclusive of those residing in the city of Calgary. The third group surveyed was a random sample of some twenty per-cent of the second-year students attending the University of Al-berta at Calgary. All of the university graduates available in the same area and holding the degrees L.L.B., B.Sc., B.A., M.D. and Ph.D. were also asked to respond to the questionnaire.

Table II summarizes the categories of subjects together with the number in each group and the percentages choosing each of the three spelling alternatives.

TABLE I
WORDS USED IN THE SPELLING SURVEY

1. aesthetic (B)	esthetic (A)	asthetic (E)
2. analyze (A)	analize (E)	analyse (B)
3. aks (E)	ax (A)	axe (B)
4. catalogue (B)	catelogue (E)	catalog (A)
5. caldron (A)	cauldron (B)	caldren (E)
6. senter (E)	centre (A)	centre (B)
7. colour (B)	coler (E)	color (A)
8. cookie (B)	cookey (E)	cooky (A)
9. distill (A)	distil (B)	disdill (E)
10. encyclopoedia (E)	encyclopaedia (B)	encyclopedia (A)
11. gipsy (B)	gypsie (E)	gypsy (A)
12. judgment (A)	judgement (B)	jugment (E)
13. moald (E)	mold (A)	mould (B)
14. programe (E)	programme (B)	program (A)
15. skilful (B)	skillful (A)	skilfull (E)
16. sulphur (B)	sulfuer (E)	sulfur (A)
17. traveler (A)	traveller (B)	travelor (E)

Legend: B—British, A—American, E—Error

Discussion

It would appear from an examination of Table II that various academic groups were differentiated on the basis of their ability to spell this selection of words correctly. However, no such conclusion can be drawn until the differences reported in this table are tested for significance. That is, the differences discovered from one group

to the next may be due to chance factors in sampling rather than to real differences in spelling ability. When the significance of the differences obtained are tested⁽⁴⁾ it is found that the Ph.D. group surpassed (obtained at >1.96) the B.Sc. group and all those groups below it. However, there is only a chance difference ($t < 1.96$) between the Ph.D.s and ranks 2, 3, and 4. Similarly, the B.A. age 20-24 years surpassed the Grade Xs and all those below them but made only chance gains over ranks 3 to 11 inclusive. The Grade Xs surpassed second-year university students but did not demonstrate significantly greater skill than any of the categories of students who had discontinued school at the end of Grade XII except the urban boys. The tendency for the city boys to spell more poorly than their rural counterparts is probably due in large measure to the fact that academically poor rural boys tend to drop out of school earlier than similar urban youth and for this reason are not included in the Grade XII group. Lack of motivation is probably an additional explanation.

TABLE II
RESULTS OBTAINED ON THE SPELLING SURVEY

Academic Level	% British Preference	% American Preference	% Errors	Number Responding	Rank Order in % Errors
Grade X Rural	53.30	39.35	7.32	1500	12
Male Grade XII Grads R*	56.70	32.30	11.00	54	15
Female Grade XII Grads. R.	55.14	39.70	5.16	57	9
Total Grade XII Grade. R.	55.89	36.10	8.01	111	13
Male Grade XII Grads. U.*	49.09	36.21	14.70	147	17
Female Grade XII Grads. U.	58.84	35.90	5.26	151	10
Total Grade XII Grades. U.	54.02	36.07	9.91	298	14
Second Year U.A.C. Students	52.39	33.30	14.31	164	16
LL.B.	64.32	32.16	3.52	62	7
B.Sc.	63.24	33.70	3.04	41	5
B.A. Male	63.32	31.33	5.35	44	11
B.A. Female	66.12	32.14	1.74	51	4
Total B.A.	64.83	31.76	3.41	96	6
**B.A. Age 20-24	69.13	29.69	1.18	32	2
**B.A. Age 40+	62.79	35.67	1.54	24	3
M.D.	62.83	32.34	4.82	31	8
Ph.D.	61.73	38.17	.19	29	1

*R.—Rural, U.—Urban
**Exclusive of B.A. male and female group above and largely female office workers.

It would seem, then, that spelling skill for the words used increases until the student enters high school and remains relatively

constant at the level throughout the three high-school years. It would also seem that spelling skill decreases when the student becomes involved in his university courses and increases again only with graduation from university.

The survey, of course, does not explain the somewhat unexpected decrement in spelling skill discovered at the second-year university level. This falling off may be due in part to the students' preoccupation with the learning of subject matter and with his need for taking copious and hurried notes during lectures. The rapid increase in vocabulary learning experienced by many students on entering university may also cause some retroactive inhibition insofar as spelling skill is concerned.

The survey demonstrates that every category responding to the questionnaire had a significant preference for the British spelling of the words used. Approximately sixty per-cent of the subjects from Grade X through second-year university preferred the British forms over the American (the summary figure of sixty per-cent is obtained by distributing the percentage of error proportionately between the British and American preferences). With graduation, the preference for the British forms increased to almost seventy per cent; (a figure arrived at by again distributing the errors).

The increase in preference for the British spelling over the American, like the decrease in skill at the second-year university level, is somewhat unexpected in the light of the preponderance of American text books and references used at the university level. The fact that the majority of university instructors obtained their terminal degrees in American Universities adds to the unexpectedness of the shift in favor of the British spelling.

The graduates responding to the questionnaire were on the average some ten years older than the undergraduates and high-school students. This fact might lead to the supposition that their teachers had a greater preference for the British spelling than those who taught the later students. However, a survey of the dictionaries, spellers, and other sources of spelling instruction in use in the Alberta schools and teachers education institutions over the past two-and-a-half decades makes this supposition untenable. The hypothesis that there is greater social prestige and acceptance attached to the older and more formal British spelling in academic circles is more tenable. The fact that the young B.A. graduate group shows the greatest preference for British spelling of any group surveyed would seem to confirm this supposition. Further testing of this hypothesis might indicate, then, that spelling preference is a social rather than an educational function.

Conclusions

The present study lacks something of rigor of design and sampling which is not wholly compensated for through the large number of subjects surveyed. Because the sample was drawn exclusively from the central area of Alberta generalization of the results is limited. The survey however, points the way for more carefully designed studies and suggests the trends that might be expected in both spelling skill and preference.

REFERENCES

- Barnhart, C. L. *The American College Dictionary*, Random House, N.Y., 1959.
- Merriam-Webster, A. *Webster's New Collegiate Dictionary*, based on Webster's New International Dictionary, 2nd Edition, Thomas Allan, Ltd., Toronto, (c) 1959.
- Fowler, H. W., Fowler, F. G. *The Concise Oxford Dictionary of Current English*, Oxford University Press, Amen House, London, 1958.
- Garrett, H. E. *Statistics in Psychology and Education*, 5th Edition Longmans, Green and Co., New York, 1958, pp. 253.

A COMPARISON OF ACHIEVEMENT OF GRADE NINE STUDENTS IN SELECTED SINGLE-GRADE AND MULTI-GRADE CLASSES IN ALBERTA

L. A. TRUCKEY AND W. D. KNILL
University of Alberta

The Problem

For purposes of school organization we have frequently assumed that academic achievement is better when there is only one grade taught in each class room. This study attempted to determine whether or not this assumption is substantiated by achievement in the departmental examination results at the grade nine level. The relationship between class organization, categorized by the number of grades per room, and student achievement, as measured by six selected grade nine departmental examinations for the school year 1961-62, was investigated.

The Selection of Schools and Classification of Students

Each year, every school in the province completes a form "A" card which is sent to the Department of Education. This card contains information such as courses taught, organization of school per classroom, number of pupils per room and number of grades per room. It was by means of these cards that schools were classified into the four groups. Once the schools were classified, the students in each group were identified. The marks achieved by each student were then obtained from the individual student cards kept by the department of education.

The basic purpose of this study was to compare the academic achievement of students in multi-grade and single grade situations. Since multi-grade situations occur only in rural areas, city and town schools were eliminated from the study. This also assisted in decreasing the socio-economic and cultural differences which may exist between schools from rural, city, and town areas. In addition, schools which were obviously different from the majority were excluded. These included separate schools, Indian schools, schools maintained by correctional agencies, and three schools which were located on the fringes of metropolitan areas. The students of the remaining schools constituted the total population which is considered to be a sample of all the students in these schools over a period of years.

Students were assigned to four categories as follows: (1) students were assigned to the single grade (SG) group if they attended schools with only one grade nine room; (2) the double-grade (DG)

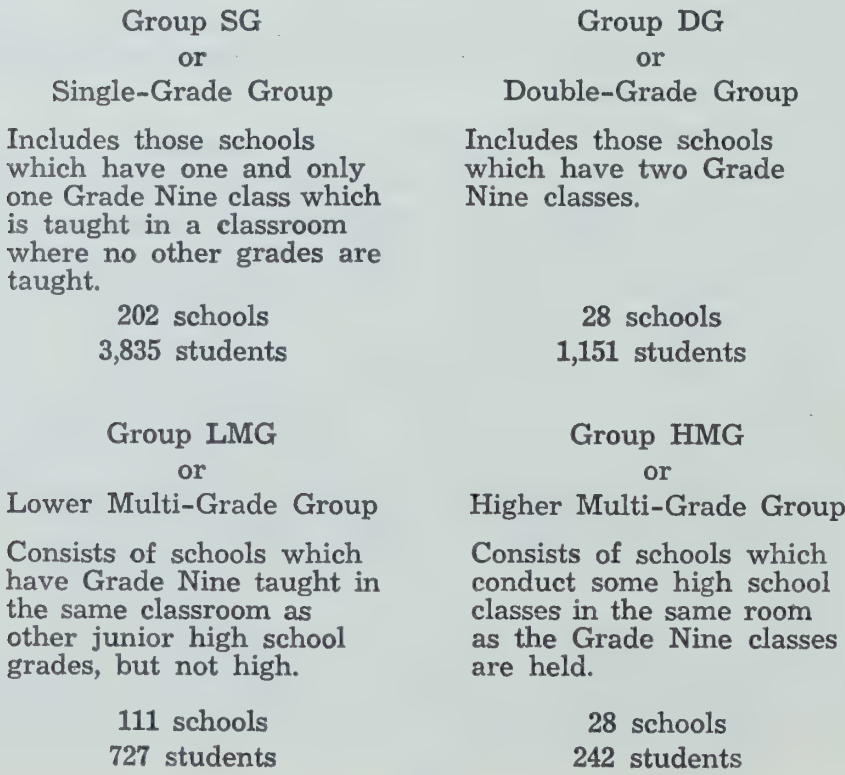
group attended schools with two grade nine rooms; (3) the lower multi-grade (LMG) group attended schools with grade nine rooms where other lower, but not higher, grades were also taught in the same room; (4) the higher multi-grade (HMG) group attended schools with grade nine rooms where some high school grades were also taught in the same room.

The single grade group (SG) consisted of 202 schools and 3,835 students, and the double grade nine group (DG) contained 28 schools and 1,151 students. The lower multi-grade (LMG) group consisted of 1111 schools and 727 students, while the higher multi-grade (HMG) group contained 28 schools and 242 students. A further description of the grouping appears in Figure 1.

TABLE I
STATUS OF THE FOUR GROUPS ON ABILITY TEST

Group	No. of Schools	Original No. of Students	Original Sum of Scores	Original Mean	Cases Cast Out	New Mean
SG Single Grade	202	3,835	260,458	67.92	1,036	63.16
DG Double Grade	28	1,151	76,335	68.06	301	63.16
LMG Lower Multi-grade	111	727	48,159	66.24	130	63.16
HMG Higher Multi-grade	28	242	15,285	63.16	Nil	63.16

FIGURE I



The Analysis

The first problem in analysis was to control for differences in ability among the four groups. This was done by equating the mean ability score of the four groups. The mean score on the Grade Nine Ability Test (SCAT) was calculated and is shown in Table I. Since the higher multi-grade group (HMG) had considerably fewer students in it, the mean ability levels of the other groups were made equal to the mean ability level of the higher multi-grade group. This was done by randomly casting out cases in the other groups, above the HMG mean, until the mean abilities were equal.

Achievement in seven areas was examined: Six examination subjects—science, mathematics, literature, reading, language and social studies were taken separately, followed by a composite achievement score of all subjects.

To determine whether differences in achievement were significant, analysis of variance was used. This involved an F-test which indicated whether or not there were significant differences among the groups. If significance was found by the F-test, it was followed by the T-test, applied to the groups taken two at a time in order to determine which specific groups were different. If the F-test did not show significance, further tests were not required. In all cases the null hypothesis was posed. Significant differences at the .10, .05, and .01 level were chosen as the criteria of significance.

The Findings

Table II summarizes the results of analysis of variance tests in each of the seven selected areas of achievement. Significant differences in mean achievement were found in four cases: composite achievement (.10 level); literature (.01 level); language (.01 level); and science (.01 level). The F obtained in analysis of variance did not indicate significant differences in reading, social studies, or mathematics.

TABLE II
SUMMARY OF ANALYSIS OF VARIANCE

Area	Significance
Composite achievement	Significant at .01 level
Reading	Not significant
Literature	Significant at .01 level
Language	Significant at .01 level
Social Studies	Not significant
Mathematics	Not significant
Science	Significant at .01 level

TABLE III
SUMMARY OF T-TESTS

Area	SG and DG	SG and LMG	SG and HMG	DG and LMG	DG and HMG	LMG and HMG
Composite	DG higher .10 level	SG higher .05 level	No significant difference	DG higher .01 level	No significant difference	No significant difference
Reading	F-test only—Assume no significant differences					
Literature	DG higher .01 level	SG higher .01 level	No significant difference	DG higher .01 level	DG higher .01 level	No significant difference
Language	DG higher .05 level	SG higher .01 level	No significant difference	DG higher .01 level	DG higher .01 level	HMG higher .05 level
Social Stud.	F-test only—Assume no significant differences					
Mathematics ...	F-test only—Assume no significant differences					
Science	DG higher .05 level	SG higher .01 level	No significant difference	DG higher .01 level	DG higher .01 level	HMG higher .01 level

Table III is a summary of the results of t-tests applied to the groups, two at a time, when analysis of variance indicated significant differences in mean achievement. The double grade (DG) group was found to have a significantly higher mean achievement than the single grade (SG) and the lower multi-grade (LMG) groups in the following areas: composite achievement, literature, language, and science. The double grade (DG) group also had a significantly higher mean achievement than the higher multi-grade (HMG) group in literature, language, and science.

The single grade (SG) group achieved significantly higher results than the lower multi-grade (LMG) group in composite achievement, literature, language, and science. No significant differences in achievement were found between the higher multi-grade (HMG) group and the single grade (SG) group. The higher multi-grade (HMG) achievement was shown to be higher than the lower multi-grade (LMG) achievement in language and science.

Generally, it may be said that the group achievements from highest to lowest were double-grade nine, single-grade nine, higher multi-grade, lower multi-grade. It should be noted that in no instance was lower multi-grade achievement found to be significantly higher than that of any other group. Equally noteworthy is the finding that double-grade achievement was the highest in every case.

Conclusion and Interpretations

Reading

Analysis of variance of reading results did not show significant differences in mean achievement. This is not too surprising for two possible reasons: First of all, the groups were equalized on ability according to the grade nine ability test; this test is largely verbal and thus, to a large degree, the groups have been equalized on reading as well as on ability. Another factor which probably affects the reading results is that reading is not taught as a subject past grade six in Alberta. It may be that by grade nine any differences which may have resulted from instruction have been overcome.

Literature

The analysis of literature results show that grade nine students from schools having two or more grade nine rooms had a significantly higher mean achievement than those students from schools where there was only one grade nine room, as well as students who received instruction in multi-grade classrooms. Students from single grade nine rooms also achieved significantly higher than did students taught in rooms where other lower grades were also taught, but not significantly higher than students in whose classrooms some high

school grades were also taught. The difference in achievement between the two classifications of multi-grade classrooms was not significant.

Language

In language achievement, students' results from schools of two grade nine rooms were significantly higher than those of the other three classifications. Single grade achievement was also significantly higher than the achievement of students taught in rooms where lower grades also received instruction, but not significantly higher than the achievement of the group which received instruction along with one or more higher grades. The multi-grade group of higher grades achieved significantly higher than did the multi-grade group having lower grades in the same room.

Social Studies

The analysis of social studies results did not produce significant differences. This is very interesting since one would expect the classroom organization and type of instruction to have a considerable effect in this subject. Social studies would seem to depend considerably upon the teacher's ability and training in that these courses in general may not be as clearly defined or follow a textbook as closely as perhaps mathematics or science. Thus a teacher would appear to be a very important factor in that a great deal of discussion and finding material from various sources would be involved. However, it may be that teachers at this level received better training for the type of instruction involved in social studies teaching than for other courses. It may also be that students in a multi-grade situation may have developed, out of necessity, the ability to work independently which appears to have some importance in grade nine social studies. Perhaps the two points just discussed may partially explain the lack of significant differences in achievement among the four groups in social studies. It would be valuable to conduct some further research into the problem of finding the most effective classroom organization for social studies teaching.

Mathematics

The analysis of mathematics results did not show significant differences in achievement. One reason may be that mathematics textbooks tend to be "the course" more than do other textbooks. Because considerable explanation, examples, tests and review exercises are contained in mathematics textbooks, it may be that students can, if they so desire, learn mathematics fairly well "on their own." The type of classroom organization and instruction may also be less effective in mathematics because of the nature of the course. Mathe-

matics at this level tends to be less nebulous and more clearly delimited than do other courses. It may be that a student who "knows the textbook" can do comparatively better in mathematics than in other subjects. Another explanation may be that our present mathematics course may stress mechanical procedures more than do other courses, and as a result directive instruction produces high achievement in examinations. Also, it may be that our mathematics teachers are using the textbook as the course and that the exams are patterned closely to the text.

Science

In science, the two grade nine room classification had a mean achievement significantly higher than did the other three groups. The students from schools where there is one complete grade nine room also achieved significantly higher results than did students from a multi-grade situation containing lower grades. The students receiving instruction in the same room with higher grades also achieved barely significantly (.10 level), higher than students who received instruction in the same room with lower grades. The differences in achievement may be due to differences in amounts of equipment supplied for instructional purposes.

Composite Achievement

The analysis of composite standings resulted in significant differences in some cases. Students from schools where there were two grade nine rooms scored significantly higher than students from schools where there was one complete grade nine room (.10 level), and also higher than the multi-grade group containing lower grades. The single grade group from schools having one complete grade nine room also showed significantly higher results than the multi-grade group with lower grades. The multi-grade group with higher grades was not significantly different from other groups.

One of the more striking findings is that although the lower multi-grade (LMG) group is found to achieve significantly lower than the other three groups, the higher multi-grade (HMG) group achievement is not found to differ significantly from the single grade (SG) achievement.

It may be that the several factors tend to increase the effectiveness of these schools as compared to the lower multi-grade group. Association with older and, presumably, more mature students may be important, as well as the "listening in" on more advanced courses in similar subject matter fields. Teachers in the higher multi-grade schools probably are in more cases better qualified to teach high school than teachers in lower multi-grade schools. This may mean that teachers in higher multi-grade schools are better qualified.

Teachers teaching grade nine in high school grades may also be more aware of the following courses and thus may emphasize "important" parts of courses to a greater extent than teachers of lower multi-grade schools.

There is also another factor which tends to make it difficult to find significant differences among various types of schools or class organizations, even though they may exist. It is likely that our ability tests not only measure natural ability but also reflect other things such as competence in reading, oral and written language, general knowledge and background of experience, to mention a few. It also seems reasonable that schools which are doing a better job of educating students would tend to produce students more capable in the areas mentioned and hence more likely to do better on the ability tests. Therefore, when we control for ability we partially nullify the effect of the school because even though the ability levels appear equal, we are possibly comparing students with higher actual ability from less effective schools to students of lower actual ability from more effective schools. The fact that the differences in mean reading achievement are extremely small tend to support this view.

Summary

The results of this study seem to justify the following generalizations:

- a. Students taught in schools where there are two grade nine rooms tend to achieve significantly higher results in grade nine departmental examinations, than do students taught in schools having one grade nine class, and lower and higher multi-grade classrooms.
- b. Students who receive instruction in schools where there is one complete grade nine room tend to achieve better results on the grade nine departmental examinations than do the students who are taught in multi-grade rooms where there are also grades lower than grade nine.
- c. Multi-grade students where higher grades are also taught tend to achieve slightly higher than multi-grade students where lower grades are also taught.
- d. This study, in general, tends to support centralization of schools on the basis of academic achievement at the grade nine level since the single grade (SG) and double-grade (DG) groups did achieve significantly higher results than did the lower multi-grade (LMG) group. However, it should be noted that on the basis of this study, grade nine students taught in the same room as higher grades achieved at approximately the same level as grade nine students taught in a single grade situation.

AN EXPERIMENTAL COMPARISON OF THE RELATIVE EFFECTIVENESS OF TWO METHODS OF TEACHING THE SOCIAL STUDIES IN GRADE ELEVEN (1960)

by
DOUGLAS WILLIAM RAY

Background And Need For The Study

The social studies teacher in Alberta is concerned with method. On the one hand the Department of Education indicates the general content and some desirable outcomes of his teaching and then leaves him free to choose suitable teaching methods. Educational theory reports the conclusions of other teachers and his own experience provides the basis for his choice.

Conant expresses the conclusions of most authorities when he writes that the three main objectives for the comprehensive high schools are

. . . first, to provide a good general education for all the future citizens; second, to provide good elective programs for those who wish to use their acquired skills immediately on graduation; third, to find satisfactory programs for those whose vocations will depend on their subsequent education in a college or university. (3)

Bining and Bining relate this view to the social studies by stating that

. . . The search for knowledge, together with its conservation and distribution is not the chief aim. Instead, the most prominent of the objectives of the secondary school is the training of pupils for effective citizenship. The materials of the social studies provide the basis for making the world of today intelligible to the pupils, for training them in certain skills and habits, and for inculcating attitudes and ideals that will enable boys and girls to take their places as efficient members of a democratic society. (2)

Andrews found that similar priorities appear to be approved by Alberta citizens of various walks of life.⁽¹⁷⁾

Comparisons between teaching methods have sometimes reflected very little concern for the development of qualities of citizenship. They have usually been concerned with measuring knowledge or understanding acquired. However, research done by Lewin and others suggests that some type of group interaction is more effective than the lecture or personal interview as a means of changing attitudes.⁽¹²⁾ The Lippitt and White investigation of autocratic, democratic and laissez-faire types of leadership revealed that pupils like the democratic system best.⁽¹³⁾ These apparently support Mofatt's contention that socializing any method makes it more effective.⁽⁸⁾ Leavitt found that highly efficient organization was often satisfactory only to the person making all the decisions.⁽¹⁾ Festinger

and Thibaut reported that group activities were especially important in shaping attitudes because of the strong desire to conform.⁽¹⁰⁾ These pressures to conform were increased when groups were closely associated and were proportional to the magnitude of deviation from the group norm. Bining and Bining believed that group pressures offered the key to the development of desirable attitudes.⁽²⁾ On the whole, these studies appear to favor teaching methods which permit a great deal of active participation of the students in the planning and execution of various activities if the development of desirable qualities of citizenship is given high priority in the educational objectives.

Principles from learning theory are also relevant to the choice of teaching method. Cronbach stressed the importance of active participation by the students in tasks such as locating and evaluating information.⁽⁴⁾ All methods of instruction provide some activity, but the two methods compared in this study differ considerably. Lecturing requires pupils to listen, think, organize and take notes. Socialized-recitation requires these activities, but in addition, information must be found, read and evaluated, and probably discussed with the teacher or a classmate. The greater apparent activity associated with socialized-recitation suggests that teaching may be more effective by this means. Horwitz recommended provision for periodic review to permit reinforcement or feedback.⁽¹⁵⁾ This requirement is probably met by the various activities associated with the socialized-recitation to a greater extent than in the lecture.

Lecturing and socialized-recitation are apparently well suited to comparison. Both methods appear to have important advantages. They have been compared by numerous procedures in different subjects. Mix compared them for social studies in Alberta (19), although the changes in the composition of the student body and organization of the school referred to by Cameron and Cheal may have made obsolete his rather cautious conclusion that lectures were efficient but not otherwise demonstrably superior.

The present experiment was an attempt to repeat the main elements of some earlier comparisons under conditions typical of a large composite high school where classes are grouped according to the previous achievement of the pupils. The comparison was made in learning of facts, development of understandings and the formation of attitudes.

The Methods Compared

The Lecture Method

The lecture method employed in the study should not be identified with the formal procedure associated with the traditional lecture.

Bining and Bining defined the lecture as any occasion on which the teacher imparted information. They said it could be short, in answer to a question, or a part of another teaching method.⁽²⁾

However, they said that lecturing was subject-matter centered, and therefore most successfully used in higher education or in those European schools which combined academically selected pupils, very well trained teachers and a disciplined atmosphere. They claimed that the lecture could be interesting or even enthralling, and that it was the teacher's job to make it so. They denied that pupils were passive during a good lecture, but were actively involved in listening, and probably summarizing or notetaking.⁽²⁾

The Socialized-Recitation

The socialized-recitation has been distinguished from the lecture in the attitude of the teacher to the learning process. The latter was not to dominate, but had to establish himself as one of the group involved in a collective effort to solve a problem.⁽²⁾

The socialized-recitation could take many forms. It is simplest, it could be a class discussion in which the teacher tried to have the class assume a major responsibility for its direction. The next progression would have a pupil as the chairman. The teacher would remain as the best informed member of the group, but would not dominate the activities with his observations or opinions. Members of the class would present information which they had previously prepared for the consideration of the group. The class would then evaluate the report in terms of the theme of the unit.⁽²⁾ The present experiment used this procedure, with each class being organized by the pupils into several committees, each of which undertook to investigate, organize and present some aspects of the unit to be discussed. The teacher's role was primarily to add significant data, to motivate, and to help in evaluating techniques so that improvement would result.

Bining and Bining claimed that socialized-recitation enabled subject matter to be better used for developing desirable attitudes and ideals.⁽²⁾ Samford and Cottle listed five advantages of the method: (1) the democratic atmosphere assures greater interest and learning; (2) leadership talent is developed; (3) the pupils cooperate to make the venture a success—class spirit results; (4) former leaders appreciate leadership problems; and (5) knowledge of correct procedure results.⁽⁹⁾

However, there is a danger that socialized-recitation might achieve the desired form but not the required spirit. If a pupil merely replaced the teacher, or if discussion flowed from a desire to please the teacher rather than from an interest in the subject

matter, the benefits could be lost.⁽²⁾ Samford and Cottle discerned five possible pitfalls: (1) pupils would be less competent than teachers as leader; (2) incorrect emphasis might result; (3) leadership talents can be developed in extracurricular activities; (4) time is wasted; and (5) teachers may feel redundant.⁽⁹⁾

It would appear that the role of the teacher is one key to the difference between the two methods. In the lecture, he would be the planner, the initiator, the controller, and the purveyor of information. In the socialized-recitation, he would be an advisor, a contriver, and a member of the group. As the role of the teacher was thereby altered, the students would have to assume a correspondingly increased responsibility for the learning process. On this difference the success of the socialized-recitation would depend.

In the experimental situation, the flexibility of method, which would normally permit a teacher to maintain interest through periodic alteration of his teaching methods, could not be permitted.

The Experimental Procedure

School Organization

The experiment was conducted with grade eleven classes at Ross Sheppard Composite High School, Edmonton. Classes were grouped into academic, general and business education programs. Within programs, students were grouped according to their grade ten average marks in academic subjects. Differences in ability for a particular subject were not always controlled but the classes were comparatively homogeneous.

The experiment was conducted in the interval between the October and Christmas examinations. Units two and five from the *Senior High School Curriculum Guide for Social Studies* 10, 20, and 30⁽¹⁶⁾ were selected for the experiment. Their subject matter was the history of industrialization and the accompanying social changes.

The Experimental Design

The experiment was conducted by using a pre-test, followed by the controlled teaching and a post-test. The rotational design advocated by Best (1) was used, so that half of the classes involved in the experiment were taught a unit by each method, and the order of the teaching methods alternated if the classes were ranked in order of their previous achievement. In this way, it was assured that each method would be used for teaching each class, for each unit, and in both first and second order. The arrangement used is shown in Table I.

TABLE I
ORGANIZATION OF THE EXPERIMENT

Classes in Rank Order	Teacher	Unit II Teaching Method	Unit V Teaching Method
<u>1</u>	A	L. ¹	S.R.
<u>3</u>	B	S.R. ²	L.
<u>5</u>	B	L.	S.R.
<u>6</u>	C	S.R.	L.
<u>9</u>	C	L.	S.R.
<u>10</u>	D	S.R.	L.
<u>14</u>	C	L.	S.R.
<u>15</u>	E	S.R.	L.

Only the underlined classes were reported in the final results.

¹ Lecture

² Socialized-Reaction

Preparation for the Experiment

The test to be used both before and after the experimental teaching was prepared from an outline of daily lesson plans for both units, and the same number of questions was prepared for each day's work. Each item could therefore be associated with one of the units, even if the lesson plans were not subsequently adhered to in every classroom. The types of questions prepared included: (1) multiple-choice, (2) matching, (3) completion, (4) short answer, and (5) attitude response. The consensus of three classroom teachers and two members of the thesis committee was used to establish (1) validity of content and (2) to separate 'fact' items from 'concepts'. Several revisions were made and unsatisfactory items were discarded.

Descriptions of the two methods of teaching, developing contrasts between them, and outlining desirable procedures were prepared in an effort to obtain uniform methods of instruction for all participating teachers. A meeting was called to discuss the experiment, to describe the methods of teaching and to answer questions by the cooperating teachers. A check list was given to each teacher so that they could report on their teaching method for each area of content. Each teacher was given a sufficient number of outlines of the content of each unit for distribution among the pupils. Each teacher had a copy of the *Curriculum Guide*.

The test was administered in two periods at the beginning and the end of the experimental teaching.

Evaluating the Results

Marking keys were prepared and the tests marked separately for each unit and for each type of question. Subjective questions

were marked from a key prepared after a number of papers had been read, and the answers were ranked so that the best ones received full marks or nearly so, and the poorest answers received few or none. Blanks and completely incorrect answers received no marks.

The attitude scores were all selected from unit five and dealt with six areas: the degree of favor expressed by the student for (1) the labor movement, (2) equalization of incomes, (3) government influence on health measures and population control, (4) preferential hiring practices, and (5) liberal and (6) socialist practices. A numerical scale was then used to study changes in the students' attitudes between the pre-test and post-test which presumably could be attributed to the teaching.

It was necessary to eliminate from the experiment the results of four of the eight classes. One teacher taught by socialized-recitation for one unit, while a practice teacher did all the lecturing for the other. The teachers did not complete the second unit on schedule. The first three and the last class listed in Table I were therefore not included. Several students' results were not included because (1) they had been observed cheating or they had been absent for one of the test periods. The effect of discarding these results was to reduce the number of students reported from nearly two hundred to sixty-eight, with no classes of superior students in the reported results.

Statistical Treatment

The formulae for significance of differences between two means for independent samples were used. These formulae were

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{(S^2/N_1) + (S^2/N_2)}}$$

where

$$S^2 = \frac{\sum (X_1 - \bar{X}_1)^2 + \sum (X_2 - \bar{X}_2)^2}{N_1 + N_2 - 2} \quad (5)$$

Conclusions

The results of the experiment are summarized in Table II. Interpretations of the significance of these results are related to the choice of the design of the experiment, the controls and the statistical procedure used. The rotational experimental design rests on four assumptions which are summarized here. (1) It must be assumed that the methods actually used by the teachers were those required by the experiment. (2) The rotational design requires the use of a post-test after the first unit as well as after the second unit.

However, if it could be shown that the combined post-test after the second unit did not favor either method, this requirement would be met. Two factors probably indicate a shortcoming in this respect: first, the two parts of the test (Unit II and Unit V) were probably not of equal difficulty; and second, the higher ability pupils (if the rank order of previous achievement also indicated the ability ranking) used the lecture method immediately before the post-test. (3) Interaction between the first method and the second must be assumed to counter-balance when the order of the teaching methods is reversed. (4) It must be assumed that the time of day when teaching or testing had no effect upon the results. The most likely factors to decrease confidence in the results of the experiment would be assumptions two and three. They are important enough to result in reservation about even the .001 significance level found in one item.

TABLE II
SUMMARY OF RESULTS OF THE EXPERIMENT

Hypothesis	Difference	Significant Level
There is no significant difference between the two methods for:		
1. presenting facts	Lecture superior	Significant
2. developing understanding	Socialized-recitation superior	Not Significant
3. influencing attitudes toward		
(a) the economic or political left	Socialized-recitation had greater effect	Not Significant
(b) sympathy toward the labor movement	Socialized-recitation had greater effect	Not Significant
(c) approval of government equalizing of incomes	Practically none	
(d) approval of government control of business	Lecture had greater effect	Not Significant
(e) approval of government influence on health, population and immigration	Socialized-recitation had greater effect	Not Significant
(f) approval of employment preferences to minimize social hardship	Practically none	

Best regarded the rotational method used in this experiment as a good procedure to minimize the uncontrolled or uncontrollable factors in classroom experiments. He claimed that the variables were controlled experimentally, rather than by statistical procedures. (1) However, other opinions indicate this choice to be undesirable. McGuigan doubted the validity of the basic assumption that the interaction between the first treatment and the second would be the same (would counter-balance) if the order of the treatments were reversed for half of the cases. (7) R. A. Fisher criticized excessive attempts to isolate various components of situations through the refinements of experiments into fragmentary units and preferred to study all factors at work simultaneously. (6)

The teaching implications of the findings are obscure. Lecturing was found to be superior as a means of imparting facts. Neither method proved significantly better for developing understanding. Socialized-recitation produced only slightly greater effect upon a pupil's attitudes than did lecturing—in this experiment by producing a slightly greater 'left-wing' tendency. There is no reason to believe that these three objectives are likely to be separated in the teaching of social studies. The superiority of either method was not tested after a longer period of time.

The comparison of the two methods upon pupils' attitudes may have been the most important part of the study. The Lickert scales were shown to be useful. However, the items used in the test would probably tend to produce relatively strong reaction, possibly exaggerating the results. There is probably no basis for concluding that a study of the history of social and political changes by either the lecture or the socialized-recitation would induce a change in students' attitudes toward either the political left or right. Other areas are probably more suited to determine the effects of various teaching methods upon student attitudes. Although none of the attitude changes found in the experiment proved to be significant, this may have been due to the relatively small number of questions in each section of the attitude tests. None of the conclusions reached in this experiment appear to have been inconsistent with the tenor of earlier studies, or with the related educational and psychological theory.

REFERENCES

A. BOOKS

1. Best, John W. *Research in Education*. Englewood-Cliffs: Prentice-Hall, Inc., 1959, pp. 135-36.
2. Bining, Arthur C., and David H. Bining. *Teaching the Social Studies in Secondary Schools*. New York: McGraw-Hill Book Company, Inc., 1952.
3. Conant, James Bryant. *The American High School Today*. New York: McGraw-Hill Book Company, Inc., 1959, p. 17.
4. Cronback, Lee J. *Educational Psychology*. New York: Harcourt Brace and Company, 1954, p. 12.

5. Ferguson, George A. *Statistical Analysis in Psychology and Education*. Toronto: McGraw-Hill Book Company, Inc., 1959, pp. 136-38.
6. Fisher, R. A. *The Design of Experiments*. New York: Hafner, 1953, pp. 91-92.
7. McGuigan, Frank J. *Experimental Psychology, A Methodological Approach*. Englewood-Cliffs: Prentice-Hall, Inc., 1960, pp. 114-15.
8. Moffat, Maurice P. *Social Studies Instruction: Organizing, Teaching and Supervising Social Studies in the Secondary Schools*. Englewood-Cliffs: Prentice-Hall, Inc., 1954, p. 150.
9. Samford, Clarence D., and Eugene Cottle. *Social Studies in the Secondary School*. New York: McGraw-Hill Book Company, Inc., 1952, pp. 221-22.

B. ARTICLES FROM COLLECTIONS

10. Festinger, Leon, and John Thibaut. "Inter-personal Communication in Small Groups," pp. 125-34, in Swanson, Guy M., et al. *Readings in Social Psychology*. New York: Henry Holt and Company, 1952.
11. Leavitt, Harold J. "Some Effects of Certain Communication Patterns on Group Performance," pp. 108-24, in Swanson, Guy M., et al. *Readings in Social Psychology*. New York: Henry Holt and Company, 1952.
12. Lewin, Kurt, in Harris, Theodore L., and Wilson E. Schwahn. *Selected Readings in the Learning Process*. New York: Oxford University Press, 1961, pp. 255-60.
13. Lippitt, Ronald, and Ralph E. White. "An Experimental Study of Leadership and Group Life," pp. 344-55, in Swanson, Guy M., et al. *Readings in Social Psychology*. New York, Henry Holt and Company, 1952.

C. PUBLICATIONS BY GOVERNMENTS OR LEARNED SOCIETIES

14. Cameron, Donald, et al. *Report of the Royal Commission on Education in Alberta*. Province of Alberta: Department of Education, 1959, p. 39.
15. National Society for the Study of Education. *The Dynamics of Instructional Groups*. Chicago: University of Chicago Press, 1960, p. 220.
16. *Senior High School Curriculum Guide for Social Studies 10, 20 and 30*. Province of Alberta: Department of Education, 1955.

D. PAMPHLETS

17. Andrews, John H. M. "Public and Professional Opinion Regarding the Tasks of the Public Schools of Alberta," *Projects in School Administration* (4) Department of Educational Supervision and Administration University of Alberta, Edmonton, 1959.

E. THESES

18. Cheal, John Ernest. "Role Conflict in the Principalship of the Composite High School." Unpublished Master's thesis, The University of Alberta, Edmonton, 1958, pp. 1-2.
19. Mix, Louis C. "An Experimental Comparison of Two Methods of Teaching Unit Two of the High School Social Studies 20 Course (1952)." Unpublished Master's thesis, The University of Alberta, Edmonton, 1954, pp. 65-70.

TEACHERS' REACTIONS TO PROCEDURES IN TEACHING HUMAN RELATIONS

METRO GULUTSAN

University of Alberta, Edmonton

Do teachers with differing attitudes toward human behavior react differently to procedures used in a program for teaching human relations? Teachers are known to differ in their views of people in ways which are significant for the educative process (Lambert, 1958; Laycock, 1950; McGee, 1955; and Ryans, 1959). It is to be expected that some of these views have a bearing on the way the teachers treat the human content of the curriculum, particularly in the area of literature and social studies. It may also be expected that teachers who vary in their conceptions of human beings and in their sensitivities to human feelings would also vary in their reactions to procedures designed to increase the effectiveness of teaching children to understand human relations.

Studies and informative descriptions dealing with the broad topic of teaching human relations have been summarized and reviewed by the Group for Preventive Psychiatry (1951), Ojemann and Clark (1956), Wall (1958), and Strang (1960). Stiles (1947), Rosenthal (1952), Mallinson (1954) and Gulutsan (1962) have investigated experimentally some effects of human relations teaching on children or on personality measures taken by children. Except for the results in Gulutsan's study, the literature on human relations teaching may be used to support the position that children's behavior and personality can be altered along directions which are regarded as improvements by those concerned with promoting public mental health.

Ojemann, director of the Preventive Psychiatry Program at the State University of Iowa, searched for a way of teaching human relations so as to extend children's insight into, and appreciation of, the causes of human behavior; to help children avoid the surface or superficial approach to behavior and to use one which is deeper and more dynamic. Ojemann (1958, p. 199) infers from the investigations undertaken by students under his direction that children can learn the beginnings of the dynamics of behavior, that they can learn to apply this knowledge in their relations with others, and that their process of learning about human behavior can be greatly extended at the school level.

Bullis and O'Malley (1947), originators of the Delaware Human Relations Classes, prepared lessons that would help children to bring

their emotional problems into the open, to realize that their problems are not unique, to examine personal problems and motivations, and to acquire insight into their own emotions. Rosenthal (1952) adapted a number of Bullis' lessons for his classroom experiment in fostering mental health among fifth grade children. He concluded, on the basis of changes in the results of sociometric and personality tests, that a teacher who employs active, interesting, purposeful group techniques can change faulty attitudes and enhance psychological functioning (Rosenthal, 1952).

Seeley (1949), working on the Forest Hill Village Project near Toronto, Canada, developed a method of teaching for mental health based on the assumption that most significant changes in behavior, and hence, in personality organization, occur in a group situation characterized by a permissive, supportive, non-punitive climate, and a structure with limits reasonable to all. In short, the method consisted of permissive, relatively unstructured discussions directed by the group itself. Mallinson (1954) found that those of the Forest Hill Village children who participated in the discussions showed larger gains on measures of personality integration than did children of the control group. The children participating in the discussions also showed larger gains in school subject grades.

Gulustan (1962) examined the effects of a human relations teaching program which included the procedures used in the studies referred to above. The instruction in human relations did have an effect, significant at the .005 level, on changes in the pupils' scores on the *Children's Manifest Anxiety Scale* developed by Castaneda, Palermo, and McCandless (1956). Manifest anxiety appeared to be more stirred up in those children who received the special instruction than it was in those who did not.

Teachers' Attitudes Toward Human Relations

The teachers' attitudes to human relations for the purposes of this study are those measured by the *Social Maturity Scale* (SM) of the *Vassar College Attitude Inventory* (Sanford, 1957). The scale consists of 149 items reported by Webster, Sanford, and Freedham (1955) as *A New Instrument for Measuring Authoritarianism in Personality*. It is made up of items which, when given to two freshman classes at Vassar College, correlated significantly with the *California F Scale* for measuring authoritarianism. A content analysis of the items showed that the SM scale measures a kind of non-authoritarianism which agrees with the theory developed in *The Authoritarian Personality* by Adorno, Frenkel-Brunswik, Levinson, and Sanford (1950). The SM scale is less obviously ideological and more personality centered and hence better able than the F scale

to measure general primitiveness (or non-primitiveness) of a subject's conception of other persons, and of his relationships with them.

Low scorers, in comparison with high scorers, possess the following traits to a larger degree: compulsiveness, punitive morality, authoritarian submission, conventionality, religious fundamentalism, anti-intraception, ego-alien symptomatology, lack of self-confidence, projectivity, cynicism, romanticism, circumscribed aggression, and a sense of victimization (Webster, 1957, p. 13-14).

Teachers who score low on the SM scale may create educational and psychological environments which are different from those provided by teachers high on the SM scale. Support for this view comes from studies in which the effects of teachers' and group leaders' personalities on pupils' behavior were examined. Lewin, Lippitt, and White (1939) made one of the earliest studies of the effects of authoritarian, democratic, and laissez-faire leadership on the behavior of children. The authoritarian leader had many of the traits possessed by low scorers on the SM scale; the democratic leader had the traits which characterize high scorers.

Anderson and Brewer (1945, 1946a, b) studied the relationship between children's behavior and the teacher's dominative or socially integrative contacts with children. Mussen summarizes the work done by Anderson and his co-workers as follows:

These studies suggest that the behavior of school children is highly dependent on the behavior of their teachers, and that teachers who use democratic (integrative) techniques with their children will be rewarded by greater cooperation, spontaneity, interest, and initiative on the part of their pupils. On the other hand, teachers who attempt to force compliance through aggressive dictatorial techniques will only be encouraging greater resistance on their pupils' part—whether manifested by whispering, doodling, lack of attention, or through more direct opposition (1956, p. 396).

Laycock (1950) assessed the effects of teachers' personality on the behavior of pupils in 157 Canadian classrooms. He prepared a list of those characteristics of teachers which he found to be associated with good mental health of pupils. It is his view that in school the greatest influence on the mental health of pupils is likely to be the mental health of the teacher (1957, p. 192). Laycock describes the teacher who has the good mental health point of view as one who, although not neglecting intellectual development, is more concerned with helping pupils learn to live; as one who regards discipline as a way of living rather than a form of control or restraint; and as one who sees the teacher in the role of a social engineer and not that of a filling station attendant.

McGee (1955) correlated teachers' overt behavior toward pupils in classrooms with the teacher's scores on the *California F-scale* and

concluded that the teachers' classroom behavior can be predicted with fair accuracy ($r = .58$) from F-scale scores. He also reported significant differences between the behavior of pupils taught by low and by high scorers. Children under the authoritarian teachers were more unresponsive, dependent, unsure, apathetic, or obstructive. In contrast, children under less authoritarian teachers were more confident, responsive, responsible, initiating, and creative.

Lambert (1958) investigated the views on educational issues held by authoritarian and nonauthoritarian teachers and principals. On the basis of statements made by them he inferred that the two contrasting types of educators advocate radically different policies on the issues of discipline and curricula.

Ryans (1959) investigated the correlates of teachers' classroom behavior in order to develop psychometric instruments for predicting this behavior. He regards the effective teacher as one who behaves in ways, and does things, that develop in pupils desirable skills, understandings, work habits, and attitudes, and adequate personal adjustment.

Morris (1955) conducted an extensive study in England on the relation of teachers' personalities to maladjustment among children. The results of his survey are in agreement with those obtained by Lewin, Anderson, Laycock, McGee, and Lambert.

The traits possessed by low scorers on the SM scale are the traits which characterize Lewin's authoritarian leader, Anderson's domineering teacher, Laycock's teacher who is a danger to the mental health of pupils, and Ryan's ineffective teacher. The high scorers, in contrast, are those likely to be described by Lewin as democratic, by Anderson as socially integrative, by Laycock as teachers with the mental health viewpoint, by McGee as equalitarian, by Lambert as nonauthoritarian, and by Ryans as effective.

The significance of teacher's social maturity and related attitudes to the teaching of mental health is apparent from the above discussion. The teacher who is low on a measure of social maturity may be expected to be little concerned with the human relations in his class. He may deal with pupils' behavior in a superficial way and not take into account the factors that underlie or cause behavior. His attitude may hinder learning in the humanities and social studies and his impersonal and less sensitive attitudes may contribute to conflicts and emotional strains among the pupils.

The teacher high on a scale measuring social maturity may be expected to show more insight into, and sensitivity for, the human relations existing in his classroom. He is likely to view pupils as human beings and not as objects which he can readily manipulate, often for his own personal satisfaction. His class control arises from

his competence in working with children rather than from his authoritative position which permits him to withhold rewards, punish or instil fear. He tries to understand the viewpoints of other pupils and does not anticipate that their behavior will necessarily follow a definite and predictable pattern. His attitude may have the following positive effects on the education and mental health of pupils:

- (a) a classroom climate, or psychological environment, which stimulates learning and promotes healthy personality development;
- (b) behavior of pupils which may be described as responsive, creative, active, and confident; and
- (c) an interaction among pupils and between pupils and teacher, sometimes described as cooperative and satisfying, which reduces social conflicts and emotional strains.

Although teachers low on SM may have more of the undesirable traits which Webster (1957, p. 14) states are characteristic of low scorers, no teacher scoring low should be expected, therefore, to have all these traits. In fact, on some traits he may be more like the high scorers. Similarly, considerable variation can be expected among high scoring teachers.

Method

Subjects

Twenty-one fourth, fifth, and sixth grade teachers, whose scores on the SM scale were known, from the public and separate schools of Moose Jaw, Canada, taught a program of human relations over a ten week period. The classes, held two times a week, were characterized by (a) unstructured, permissive group discussions patterned on those employed in the Forest Hill Village experiment (Seeely, 1954); (b) catharsis or the expression of troubled feelings as suggested by some psychiatrists (e.g. Ross, 1957); (c) a psychodynamic emphasis of the human content in the stimulus materials used in the curriculum as recommended by Ojemann (1949); and (d) role playing as adapted for classroom use by Jennings (1950). The investigator familiarized the teachers with these procedures during three ninety minute meetings attended by all the teachers, by classroom demonstrations, by individual conferences, and by means of a guide² which included a set of readings.

Reaction Questionnaire

Teachers reactions to the procedures were obtained during a personal interview six weeks after they had completed the ten weeks of instruction described above. The teachers were asked the following questions:

2. Available from author.

1. Rank in order of your preference the four instructional procedures used in the mental health (human relations) education experiment.
2. Which procedures, if any, did you use before the experiment began?
3. Which procedures, if any, did you use in teaching literature, social studies, and subjects other than health (the one to which the experimental procedures were attached)?
4. Which procedures do you expect to use regularly, which occasionally, and which not at all?
5. Which procedures appeared to have a favourable influence on some pupils?

Teachers were also asked to offer additional observations and comments about the program in which they had participated. The reactions of teachers whose scores on the SM scale were, in respect to the group of teachers in this study, in the top third compared with the reactions of teachers whose scores were in the lower third.

Results

The results obtained from teachers are summarized in Table I and the statistical analyses of the replies are reported in Tables 2 and 3. The reader, after even a quick inspection of the rankings in question 1 summarized in Table 1, may be struck by the decidedly contrasting reactions of the teachers to the procedures. The average rankings of (a) unstructured, permissive group discussions, (b) expressing feelings, a form of catharsis, (c) stimulus materials in which the human content of instructional materials is given a psychodynamic emphasis, and (d) role playing given by the high scoring teachers were 1.4, 2.1, 3.0, and 3.6, and by the low scoring teachers 3.1, 2.6, 2.4, and 1.9 respectively. The average ranking of the two groups of teachers were in reverse order, making the rank order correlation -1.00. It is true that this correlation was based on four pairs of average rankings; but it must be noted that each ranking is based on seven rankings.

Kendall (1948) has proposed a coefficient of concordance among rankings and has obtained the sampling distribution of this statistic. The degree of concordance between the rankings by the high and low scoring teachers was very low, as may be seen in Table 2. Among the seven high scoring teachers the coefficient of concordance was significant at the .01 level; among the seven low scoring teachers agreement was not significant. Because tables for evaluating discordance are not available, the rankings of the low scoring group were reversed, i.e., a ranking of 4 was taken as 1, and 1 was taken as 4, and the coefficient of concordance was calculated; the agreement among the total group of fourteen teachers was found to be significant at the .05 level. Therefore, the writer concluded that rankings by the two groups were significantly discordant, discordance being defined as concordance between two sets of rankings, one set of which has been reversed.

TABLE I
TEACHERS' ANSWERS TO QUESTIONS ASKED
SIX WEEKS AFTER THE EXPERIMENT

Teacher	Question 1				Question 2				Question 3				Question 4				Question 5			
	UD	EF	SM	RP	UD	EF	SM	RP	UD	EF	SM	RP	UD	EF	SM	RP	UD	EF	SM	RP
1	1	2	3	4	x	x	x	x	x	x	x	x	r	r	r	r	x	x	x	-
2	1	2	3	4	x	x	x	x	x	x	x	x	r	r	r	r	x	x	-	-
3	4	2	1	3	x	x	x	x	x	x	x	x	-	o	o	o	-	-	x	x
4	1	3	2	4	x	x	x	-	-	-	x	x	r	-	o	o	x	x	x	x
5	1	2	3	4	x	-	x	-	x	x	x	x	o	o	o	o	x	x	x	x
6	1	2	3	4	x	x	x	-	x	-	x	x	r	r	r	r	x	x	x	x
7	1	2	3	4	x	x	x	-	x	x	-	-	o	r	o	r	x	x	-	x
8	1	2	3	4	x	x	x	-	x	-	-	-	o	o	o	o	x	-	-	-
9	1	2	3	4	x	-	x	x	x	-	-	x	o	o	o	o	x	x	x	-
10	2	3	4	1					x	x	-	x	o	o	o	o	x	-	-	x
11	2	3	4	1	-	-	x	-	x	-	-	-	o	-	o	o	x	-	x	x
12	1	4	3	2	x	x	x	-	x	x	x	-	r	r	r	r	x	x	-	x
13																				
14	1	2	3	4	-	-	x	-	x	-	x	x	r	o	o	o	x	-	x	-
15	4	2	1	3	x	x	x	-	-	-	x	x	o	o	o	o	x	x	x	x
16	1	3	4	2	-	x	-	-	-	-	-	x	r	r	-	r	x	-	-	x
17	3	1	4	2					x	x	x	-	o	o	o	o	-	x	-	-
18	2	4	1	3	-	-	x	-	-	-	-	-	-	-	-	-	-	-	-	-
19	4	3	2	1					-	-	-	x	-	-	o	o	-	-	-	-
20	4	2	3	1	-	-	-	-	-	-	-	-	o	o	-	r	-	-	-	x
21	4	3	2	1	-	-	-	-	-	-	-	-	-	o	-	o	-	-	-	-

Notes:— Teachers identified by numbers 1 to 7 were among the upper third on the SM scale; teachers 8 to 14 in the middle third; and teachers 15 to 21 were in lowest third. There are no entries for number 13 because the teacher was not available for interviewing. Teachers 10, 17 and 19 were new to the profession and were not able, therefore, to answer Question 2.

Teachers' answers to Question 1 are in terms of ranks, to Question 2, 3, and 5 in terms of an x, if affirmative, and -, if negative, to Question 4 in terms of r, if regular, and o, if occasional, and -, if neither.

The instructional procedures are abbreviated as follows:

- UD—Unstructured, permissive group discussions;
- EF—Expressing feelings, a form of catharsis;
- SM—Stimulus materials in which human relations are given dynamic emphasis;
- RP—Role-playing, a form of sociodrama or psychodrama.

TABLE 2
ANALYSIS OF RANKS ASSIGNED TO THE FOUR TYPES OF INSTRUCTION BY HIGH AND LOW SCORING TEACHERS

Group	No. of Teach-ers	Observed S values ^a	S value at	
			.05 level	.01 level
High scoring teachers	7	125	88.7	118
Low scoring teachers	7	41	88.7	118
Combined group	14	26	180	251
Combined group after reversing ranks of low group	14	226	186	251

^a—S values are used in the same sense as by Walker and Lev (1953, p. 284).

TABLE 3
FREQUENCY OF AFFIRMATIVE RESPONSES GIVEN TO QUESTIONS 2, 3, 4, and 5 BY TEACHERS HIGH (H) AND LOW (L) ON THE SOCIAL MATURITY SCALE

Frequency of Affirmative Responses	Question 2		Question 3		Question 4		Question 5	
	H	L	H	L	H	L	H	L
4	3	...	4	...	5	2	3	1
3	3	1	1	1	2	2	2	...
2	1	...	2	1	...	2	2	1
1	...	2	...	2	...	...	...	2
0	...	2	...	3	...	1	...	3
KD	.800*		5†		3†		5†	
KD significant at .05 level‡	Yes		Yes		No		Yes	

* As there are more high scorers than low scorers answering question 2, the difference is given in decimal equivalents and Siegel's formula 6.11 (1956, p. 131) was used for computing the Chi-square value for the Kolmogrov-Smirnov two-sample test in this question.
† Siegel (1956, p. 278) presents a table of critical values of KD-cumulative deviation.
‡ The one-tailed test is used because the writer is testing the significance of the deviation in a predicted direction.

Differences may be observed in the frequency of affirmative responses given by high and low scoring teachers on questions 2, 3, and 5 (see Table 3). It was possible to offer up to four affirmative responses, one for each procedure, in each of the questions asked.

High scoring teachers reported much more frequently than did low scoring teachers that they were using the special procedures before the beginning of the experiment, and are using them in teaching literature, social studies, and other subjects. They also reported significantly more frequently than did low scorers that the effects of the procedures on pupils were favourable. No differences were found between the types of teachers in respect to expected uses of the procedures. It must be noted, however, that teachers were permitted to answer question 4 with *occasional*, which was treated as affirmative in the analysis but which may only have provided teachers with an opportunity to be noncommittal.

At the six-week follow-up interview, teachers were asked to offer additional observations about the program in which they had participated. The comments are listed below with the teacher's identifying number from Table 1.

Teachers' Comments

1. Pupils prefer role-playing even though I think less of it than of the other procedures.
2. Talking things out and expressing feelings may be more effective than the unstructured discussion.
3. My class did not respond to the unstructured discussion.
4. The general cumulative effect seems good.
5. I like the variety that I find in the procedure.
10. The procedures gave me a key to understanding pupils' behavior.
11. The procedures are fine for teaching health.
14. I was able to employ role-playing for coping with a situation affecting class discipline.
15. Children liked role-playing best.
16. The free discussions influenced some children a great deal. The children talk things over more readily now.
17. I noticed that the program had some influence on my teaching. It helped me to organize other teaching material.
18. I now use a workbook in health, but children still want to role-play.
19. From an over-all point of view the program had some influence on children.
20. I did not find enough variety in the procedures. They were very repetitious.

Teachers with differing attitudes, as measured by the SM scale, commented differently on the procedures in the mental health program. The low-scorers reiterated their preference for the role-playing procedure, the high scorers for free discussions and talking things out or expressing feelings. One high scoring teacher remarked on the variety she found in the procedures; a low scoring teacher remarked on the lack of variety. On the basis of these comments and the responses reported in Table 1, it may be inferred that teachers with differing attitudes as revealed by the SM scale do react

differently to the instructional procedures recommended by Seeley, Ojemann, and others for teaching in the area of human relations and mental health.

Additional support for this inference comes from the teachers' reactions gathered one year after the experiment (Gulutsan, 1962, pp. 90-99). Teachers high on the SM scale still regarded the program favorably. Teachers low on SM still regarded it unfavorably but not as unfavorably as during the earlier interviews. Again, the low scorers remarked on the lack of variety in the program and that they preferred the role-playing procedures that were part of the experiment.

Discussion and Conclusion

The significant variation of reactions to the procedures used in the teaching of human relations which was found between teachers scoring high and those scoring low on the SM scale could have been expected from the theory underlying the authoritarian personality. The rationale for expecting differences between low and high authoritarian teachers is based on the work done by Adorno and associates (1950), Else Frenkel-Brunswik (1949), Sanford (1956), and Rokeach (1960). One could expect teachers with high SM scores (low on authoritarianism) to be comfortable with the minimally structured group discussions; on the other hand, one could expect teachers low on SM to prefer role playing where the degree of structure is under more direct control of the teacher. Thus, the preference for the degree of structure in instructional procedures for extending children's understanding of human relations appears to be a function of the teacher's personality. This preference may have a significant bearing, probably superseding that of course content, on what children learn about people from their teachers.

Summary

Twenty-one fourth, fifth and sixth grade teachers whose scores on the Social Maturity (SM) Scale of the Vassar College Attitude Inventory were known, taught a program of human relations in which they were required to use (a) unstructured, permissive group discussions, (b) expression of troubled feelings, (c) discussion of materials in which human behavior was given a dynamic emphasis, and (d) role playing. Teachers high on SM tended to rank the procedures in the order listed above; those low on SM ranked them in the reverse order. The rankings were significantly discordant ($p < .05$). High scorers reported more frequent transfer of the procedures to other subject areas than did low scorers.

REFERENCES

- Adorno, T. W., Frenkel-Brunswick, E., Levinson, D. J. and Sanford, R. N. *The authoritarian personality*. New York: Harper, 1950.
- Anderson, H. H., and Brewer, J. E. Studies of teachers' classroom personalities, I: Dominative and socially integrative behavior of kindergarten teachers. *Applied Psychology Monographs*, No. 6, 1945.
- , Studies of teachers' classroom personalities, II: Effects of teachers' dominative and integrative contacts on children's classroom behavior. *Applied Psychology Monographs*, No. 8, 1946a.
- , Studies of teachers' classroom personalities, III: Follow-up studies of the effects of dominative and integrative contacts on children's behavior. *Applied Psychology Monographs*, No. 11, 1946b.
- Bullie, H. E. and O'Malley, E. E. *Human relations in the classroom, course I*. Wilmington, Delaware: Hambleton, 1947.
- Castaneda, A., McCandless, B. R. and Palermo, D. S. The Children's form of the manifest anxiety scale. *Child Development*, 1956, 27, 317-326.
- Jennings, H. H. Sociodrama as educative process, in Association for Supervision and Curriculum Development, 1950 Yearbook. *Fostering mental health in our schools*, 1950, 260-285.
- Kendall, M. G. *Rank correlation methods*. London: Charles Griffin, 1948.
- Lambert, P. Interaction between authoritarian and nonauthoritarian principals and teachers. *Genetic Psychology Monographs*, 1958, 58, 163-205.
- Laycock, S. R. Effects of the teacher's personality on the behavior of pupils. *Understanding the child*, 1950, 19, 50-55.
- Laycock, S. R. The teacher's influence on the mental health of pupils, *The School, Secondary Edition*, 1957, 26, 191-195.
- Lewin, K., Lippitt, R. and White, R. K. Patterns of aggressive behavior in experimentally created "social climates." *Journal of Social Psychology*, 1939, 10, 271-299.
- Mallinson, T. J. *An experimental investigation of group directed discussion in the classroom*. Unpublished Ph.D. dissertation, University of Toronto, 1954.
- McGee, H. M. Measurement of authoritarianism and its relation to teachers' classroom behavior. *Genetic Psychology Monographs*, 1955, 52, 89-146.
- Morris, B. S. Mental Hygiene in the classroom—teacher's personality and problems of children's adjustment. In London University Institute of Education: *Bearing of recent advances in psychology of educational problems*. London: Evans. 1955.
- Mussen, P. H. *Child development and personality*. New York: Harper, 1956.
- Ojemann, R. H. Integrated plan for education in human relations, in American Educational Research Association, *Growing points in educational research*, 1949b, 213-218.
- . Basic approaches to mental health: The human relations programme at the State University of Iowa. *Personnel and Guidance Journal*, 1958, 37, 199-206.
- Ojemann, R. H. and Clark, W. J. School—community programmes. *Review of Educational Research*, 1956, 26, 479-501.
- Rokeach, M. *Open and closed minds*. New York: Basic Books, 1960.
- Rosenthal, S. A fifth grade classroom experiment in fostering mental health. *Journal of Child Psychiatry*, 1952, 2, 302-329.
- Ross, Helen. When in trouble, talk it out. *Just like me, teacher's edition*. Chicago: Scott-Foresman, 1957, p. 77.
- Ryans, D. G. Some correlates of teacher behavior. *Educational and Psychological Measurements*. 1959, 19, 3-12.

- Sanford, R. N. The approach to the authoritarian personality. In McCary, J. L. (Ed.) *Psychology of Personality*. New York: Logos Press, 1956.
- Sanford, N. R. *Vassar College Attitude Inventory*. Poughkeepsie, New York: Harmon Printing, 1957.
- Seeley, J. R. Come let us reason together. *Forest Hill Village Home and School Post*. March, 1949.
- Seeley, J. R. The Forest Hill Village experiment. *Understanding the Child*, 1954, 23, 104-110.
- Siegel, S. *Nonparametric statistics*. New York: McGraw-Hill, 1956.
- Stiles, F. S. *A study of materials and programmes for developing an understanding of behavior at the elementary school level*. Unpublished Ph.D. dissertation, State University of Iowa, 1947.
- Strang, R. M. Mental health. In Harris, C., (Ed.) *Encyclopedia of Educational Research*, third edition. New York: Macmillan, 1960.
- Walker, H. H. and Lev, J. *Statistical inference*. New York: Holt, 1955.
- Wall, W. D. Curriculum and the promotion of mental health: and historical and comparative review. *The Yearbook of Education*, London: Evans Brothers, 1958.
- Webster, H. *Research manuals VC Attitude Inventory and VC Figure Preference Test*. Poughkeepsie, New York: Mary Conover Mellion Foundation, 1957.
- Webster, H., Sanford, N. R. and Freedman, M. A new instrument for studying authoritarianism in personality. *The Journal of Psychology*, 1955, 40, 73-84.

BOOK NOTICES

Cohen, Alan, and Garner, Norman. *A Student's Guide to Teaching Practice*. Clarke, Irwin & Co., Toronto, 1963. pp. 121, \$1.80.

A text for students in training colleges which deals with problems raised in practice teaching. It considers in detail the preliminary visit, the preparation of lessons, the relationship between the student and the teacher and ways of evaluating the students' work.

Thyne, James. *The Psychology of Learning and Techniques of Teaching*. Clarke, Irwin & Co., Toronto, 1963. Pp. 236, \$3.80.

In this book the author works out in detail the practical applications of the psychology of learning to the actions the teacher takes in the classroom.

Hamlin, D. L. B., and Lalande, Gilles. *International Studies in Canadian Universities*. Canadian Universities Foundation, 75 Albert, Ottawa, 1964. Pp. 220.

A survey of course offerings and facilities for undergraduate and graduate students in international studies. The fields covered are Russian and European, Asian, African, and Latin American studies. A chapter is devoted to needs in this area of study, and a final chapter to the international role of Canadian Universities. Half of the book is written in French by the French-language editor.

The Editorial Committee for the
ALBERTA JOURNAL OF EDUCATIONAL RESEARCH

will be pleased to consider research articles from contributors outside the Province of Alberta, for publication in the Journal.

Manuscripts should be addressed to the Editor.

H. E. Smith

